ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES

MSc THESIS

Umut Ahmet ÇETİN

PREDICTING COVID-19 INFECTION USING MACHINE LEARNING AND FEATURE SELECTION METHODS

DEPARTMENT OF COMPUTER ENGINEERING

ADANA, 2022

ABSTRACT

MSc THESIS

PREDICTING COVID-19 INFECTION USING MACHINE LEARNING AND FEATURE SELECTION METHODS

Umut Ahmet ÇETİN

ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES
DEPARTMENT OF COMPUTER ENGINEERING

Supervisor : Assoc. Prof. Dr. Fatih ABUT
Year: 2022, Pages: 199
Jury : Assoc. Prof. Dr. Fatih ABUT
: Prof. Dr. M. Fatih AKAY
: Assoc. Prof. Dr. Fatih KILIÇ

The purpose of this thesis is to develop new COVID-19 prediction models using different machine learning and feature selection methods. Particularly, Multi-Layer Perceptron (MLP), Tree Boost (TB), Radial Basis Function Network (RBF), Support Vector Machine (SVM), and K-Means Clustering (KMC) have been used to construct various regular and feature selection-based COVID-19 prediction models. The minimum redundancy maximum relevance (mRMR) and Relief-F algorithms have been chosen as the feature selectors. The dataset has information related to 20.000 patients (i.e., 10.000 positives, 10.000 negatives) and includes several personal, symptomatic, and non-symptomatic variables. The accuracy, precision, recall, and F1-score metrics have been used to assess the models' performance, whereas the generalization errors of the models has been evaluated using 10-fold cross-validation. The results show that, in general, MLP outperforms all other ML classifiers for predicting the COVID-19 infection. The average performance of mRMR is slightly better than Relief-F in predicting the COVID-19 infection of a patient. The symptom-based variables such as fever, cough, and headache have been found as the most vital predictors of COVID-19 infection.

Key Words: COVID-19, coronavirus, machine learning, feature selection, prediction

ÖZ

YÜKSEK LİSANS TEZİ

MAKİNE ÖĞRENMESİ VE NİTELİK SEÇME YÖNTEMLERİYLE COVID-19 TAHMİNİ

Umut Ahmet ÇETİN

ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

Danışman : Doç. Dr. Fatih ABUT
Yıl: 2022, Sayfa: 199
Jüri : Doç. Dr. Fatih ABUT
: Prof. Dr. M. Fatih AKAY
: Doç. Dr. Fatih KILIÇ

Bu tezin amacı, çeşitli makine öğrenme ve nitelik seçme yöntemlerini kullanarak yeni COVID-19 tahmin modelleri geliştirmektir. Çok Katmanlı Algılayıcı (MLP), Tree Boost (TB), Radyal Temel İşlev Ağı (RBF), Destek Vektör Makinesi (SVM) ve K-Ortalama Kümeleme (KMC) yöntemleri kullanılarak çeşitli regüler ve nitelik seçme tabanlı COVID-19 tahmin modelleri oluşturmuştur. Minimum fazlalık maksimum önem (mRMR) ve Relief-F algoritmaları nitelik seçme yöntemleri olarak seçilmiştir. Veri seti, 20.000 hasta (10.000 pozitif, 10.000 negatif) ile ilgili bilgileri içermektedir ve çeşitli kişisel, semptomatik ve asemptomatik değişkenlerden oluşmaktadır. Modellerin performansını değerlendirmek için doğruluk, kesinlik, hassaslık ve F1-skor metrikleri kullanılırken, modellerin genelleme hataları 10-katlı çapraz doğrulama kullanılarak değerlendirilmiştir. Sonuçlar, genel olarak MLP'nin COVID-19 enfeksiyonunu tahmin etmede diğer tüm ML sınıflandırıcılarından daha iyi performans sergilediğini göstermektedir. mRMR'nin ortalama performansının, bir hastanın COVID-19 enfeksiyonunu tahmin etmede Relief-F'e göre biraz daha iyi olduğu gözlemlenmiştir. Ateş, öksürük ve baş ağrısı gibi semptom temelli değişkenler, COVID-19 enfeksiyonunun en önemli belirleyicileri olarak bulunmuştur.

Anahtar Kelimeler: COVID-19, koronavirüs, makine öğrenmesi, nitelik seçme, tahmin

EXTENDED ABSTRACT

COVID-19 is an infection that has affected the world since December 31, 2019, and was declared a pandemic by WHO in March 2020. The COVID-19 pandemic has infected 465 million people and claimed more than 6 million lives. It is also known as Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). SARS-CoV-2 causes severe pneumonia along with a fatality rate of 2.9%. Patients infected with SARS-CoV-2 may have asymptomatic, mild, or severe symptoms. The most common symptoms of SARS-CoV-2 include cough, shortness of breath, and fever but also symptoms like vomiting and diarrhea. SARS-CoV-2's transmission is mainly through the respiratory route like other respiratory viruses. Patients suspected of having SARS-CoV-2 are verified mostly through reverse real-time PCR (rRT-PCR) tests.

There are already many studies about disease predictions, many studies already conducted about diseases like swine flu, h1n1 flu, and influenza. However, characteristics and spread rate of COVID-19 make it different from other diseases. It forced many governments and businesses to take many harsh measures and caused people to develop new habits. That's why developing new models to predict this disease is important. The development of new prediction models may help governments and businesses to limit the spread of COVID-19, arranging adequate amount of supply and personal, or planning future decisions so they can avoid many problems in the future such as insufficient personal and supplies, or overfilled facilities.

The purpose of this thesis is to develop new COVID-19 prediction models. Two types of prediction models have been developed, namely regular prediction models and feature selection-based prediction models. Regular COVID-19 prediction models were created using the predictor variables' unique, double, triple, quadruple, quintuple, sextuple, septuple, and octuple combinations. Moreover, feature selection-based models were also created using two state-of-art feature

selection methods including minimum redundancy maximum relevance (mRMR) and Relief-F. To create prediction models, five different ML including Multi-Layer Perceptron (MLP), Tree Boost (TB), Radial Basis Function Networks (RBF), Support Vector Machines (SVM), and K-Means Clustering (KMC) have been used for model development.

The dataset used in this study was obtained from the Espirito Santo State Portal of Brazil on May 23, 2021. The dataset contains over 40 columns and over 1.4 million rows. The dataset contains data that was collected between March 17, 2020, and May 23, 2021, and includes a diverse amount of knowledge. The common predictor variables appearing in each dataset are age, sex, fever, breathing difficulty, cough, runny nose, throat pain, diarrhea, and headache. In addition to these variables, the dataset also contains race, pregnant, lung comorbidity, cardio comorbidity, renal comorbidity, diabetes comorbidity, smoking comorbidity, and obesity comorbidity.

By utilizing the mRMR and Relief-F feature selectors, the rank of each variable has been calculated and sorted in descending order by their ranks. Then, each variable with the lowest rank were successively removed. By doing so, 17 mRMR-based and 17 Relief-F-based prediction models were created. The performance of each model has been evaluated using 10-fold cross-validation and calculating the accuracy, recall, and F1-score values of the models.

The results reveal that unique, double, triple, quadruple, quintuple, sextuple, septuple, and octuple combinations of the predictor variables do not have any significant performance difference when the performance of the models is compared. Furthermore, it is also seen that feature selection algorithms also do not provide any significant performance improvement to the models. The results also show that using comorbidities such as obesity, diabetes, smoking, or renal comorbidity, or using biological characteristics such as age, gender, or race does not improve the performance of the results significantly.

The results show that the average performance of mRMR is slightly better than Relief-F. Furthermore, mRMR is more successful than the Relief-F algorithm

in finding the relative relevance order of the predictor variables to predict COVID-19 infection. Finally, the mRMR feature selector emphasizes symptomatic variables such as fever and cough, whereas the Relief-F algorithm emphasizes non-symptomatic variables such as age and race. Also, when the average results of feature selection-based models are compared with other models, it is seen that those two types of models do not have any superiority among each other.

When the accuracy, recall, and F1-score values of the models are compared, it is seen that the MLP-based models, in general, exhibit the best performance. Except for recall, the performance of TB-based, RBF-based, and SVM-based models are similar/comparable. In other words, there is no superiority of one classifier over the other two classifiers, but their accuracy, recall, and F1-score values are lower than those of MLP-based models. Finally, the KMC-based models produce the lowest accuracy, precision, recall, and F1-score values. In contrast, when the recall values of the models are compared, the general order of ML classifiers leading from the best to the worst results is MLP, TB, RBF, SVM, and KMC. However, when the average precision values of the models are compared, it is seen that TB and SVM performed best in this metric, and also that the performance of TB and SVM were similar/comparable. When the average precision values of the RBF, MLP, and KMC are compared, it is seen that the general order of ML classifiers leading from the best to the worst results is RBF, MLP, and KMC.

Lastly, for the importance of the variables for COVID-19 prediction, the symptomatic variables such as fever, cough, and headache have been found as the most vital variables for predicting the COVID-19 infection. On the other hand, non-symptomatic variables such as age, sex, race, and obesity comorbidity have a negligible effect on the performance of the models.

GENİŞLETİLMİŞ ÖZET

COVID-19, 31 Aralık 2019'dan beri dünyayı etkileyen ve Mart 2020'de DSÖ tarafından pandemi ilan edilen bir enfeksiyondur. COVID-19 pandemisi 465 milyon kişiye bulaşmıştır ve 6 milyondan fazla can almıştır. Şiddetli Akut Solunum Sendromu Coronavirüs 2 (SARS-CoV-2) olarak da bilinmektedir. SARS-CoV-2, %2.9'luk bir ölüm oranıyla birlikte ciddi pnömoniye neden olmaktadır. SARS-CoV-2 ile enfekte olan hastalarda asemptomatik, hafif veya şiddetli semptomlar olabilmektedir. SARS-CoV-2'nin en yaygın semptomları arasında öksürük, nefes darlığı ve ateşin yanı sıra kusma ve ishal gibi semptomlar da bulunmaktadır. SARS-CoV-2'nin bulaşması, diğer solunum yolu virüsleri gibi esas olarak solunum yolu yoluyla gerçekleşmektedir. SARS-CoV-2 olduğundan şüphelenilen hastalar çoğunlukla ters gerçek zamanlı PCR (rRT-PCR) testleri ile doğrulanmaktadır.

Literatürde domuz gribi, h1n1 gribi, grip vb. hastalıklarla ilgili halihazırda yapılmış birçok tahmin çalışması bulunmaktadır. Ancak COVID-19'un özellikleri ve yayılma hızı onu diğer hastalıklardan farklı kılmaktadır. Bu hastalığın farklı özellikleri ve yayılma hızı, birçok hükümeti ve işletmeyi sert önlemler almaya zorlamıştır ve insanların yeni alışkanlıklar geliştirmesine sebep olmuştur. Bu nedenle bu hastalığı öngörmek için yeni modeller geliştirmek önemlidir. Yeni tahmin modellerinin geliştirilmesi, hükümetlerin ve işletmelerin COVID-19'un yayılmasını sınırlamasına, yeterli miktarda tedarik ve personel düzenlemesine veya gelecekteki kararları planlamalarına yardımcı olabilir. Böylece gelecekte yetersiz personel ve sarf malzemeleri veya tesislerde aşırı dolum gibi birçok sorundan kaçınabilirler.

Bu tezin amacı, yeni COVID-19 tahmin modelleri geliştirmektir. Düzenli ve nitelik seçme tabanlı tahmin modelleri olmak üzere iki tür tahmin modeli geliştirilmiştir. Tahmin değişkenlerinin birli, ikili, üçlü, dörtlü, beşli, altılı, yedili ve sekizli kombinasyonları kullanılarak düzenli COVID-19 tahmin modelleri oluşturulmuştur. Ayrıca, Relief-F ve mRMR ile bir ranker arama yöntemi

kullanılarak nitelik seçme tabanlı modeller de oluşturulmuştur. Tahmin modellerini geliştirmek için, Çok Katmanlı Algılayıcı (MLP), Tree Boost (TB), Radyal Temel İşlev Ağları (RBF), Destek Vektör Makineleri (SVM) ve K-Means Kümeleme (KMC) olmak üzere beş farklı ML kullanılmıştır.

Bu çalışmada kullanılan veri seti, 23 Mayıs 2021 tarihinde Brezilya'nın Espirito Santo Eyalet Portalı'ndan elde edilmiştir. Veri seti 40'tan fazla sütun ve 1.4 milyondan fazla satıra sahiptir. Veri kümesi, 17 Mart 2020 ile 23 Mayıs 2021 arasında toplanan verileri içermektedir. Yaş, cinsiyet, ateş, nefes alma zorluğu, öksürük, burun akıntısı, boğaz ağrısı, ishal ve baş ağrısı değişkenleri, tüm düzenli COVID-19 tahmin modellerinde ortak olarak kullanılmış değişkenlerdir. Bu değişkenlere ek olarak ırk, hamilelik, akciğer komorbiditesi, kardiyo komorbiditesi, renal komorbidite, diyabet komorbiditesi, sigara komorbiditesi ve obezite komorbiditesi gibi değişkenler de veri setinde bulunmaktadır.

mRMR ve Relief-F nitelik seçicileri kullanılarak her bir değişkenin sırası hesaplanmış ve sıralarına göre azalan düzende sıralanmıştır. Ardından, model geliştirme aşamasında en düşük sıradaki her bir değişken art arda çıkarılmıştır. Bu şekilde 17 mRMR tabanlı ve 17 Relief-F tabanlı tahmin modeli oluşturulmuştur. Her bir modelin performansı, 10-kat çapraz doğrulama kullanılarak ve modellerin doğruluk, hassaslık ve F1-skoru değerleri hesaplanarak değerlendirilmiştir.

Sonuçlar, tahmin değişkenlerin birli, ikili, üçlü, dörtlü, beşli, altılı, yedili ve sekizli kombinasyonlu modellerin performansları karşılaştırıldığında anlamlı bir performans farkının olmadığını ortaya koymaktadır, ayrıca nitelik seçme yöntemlerinin modellerde önemli bir performans artışı sağlamadığı da görülmüştür. Sonuçlar ayrıca obezite, diyabet, sigara içme veya renal komorbidite gibi komorbiditelerin veya yaş, cinsiyet veya ırk gibi biyolojik özelliklerin kullanılmasının sonuçların performansını önemli ölçüde iyileştirmediğini göstermektedir.

Sonuçlar, mRMR'nin ortalama performansının Relief-F'den biraz daha iyi olduğunu göstermektedir. Ayrıca, mRMR, COVID-19 enfeksiyonunu tahmin etmek

için tahmin değişkenlerin göreceli önem sırasını bulmada Relief-F algoritmasından daha başarılıdır. Son olarak, mRMR nitelik seçicisi ateş ve öksürük gibi semptomatik değişkenlere daha fazla önem verirken, Relief-F algoritması yaş ve ırk gibi semptomatik olmayan değişkenlere daha çok önem vermiştir. Ayrıca nitelik seçimine dayalı modellerin ortalama sonuçları diğer modellerle karşılaştırıldığında bu iki tür modelin birbirlerine üstünlüklerinin olmadığı görülmektedir.

Modellerin doğruluk, duyarlılık ve F1-skor değerleri karşılaştırıldığında, genel olarak MLP tabanlı modellerin en iyi performansı sergilediği görülmektedir. Hassaslık, TB tabanlı, RBF tabanlı ve SVM tabanlı modellerin performansı benzer/karşılaştırılabilir. Diğer bir deyişle, bir sınıflandırıcının diğer iki sınıflandırıcıya üstünlüğü yoktur, ancak doğruluk, duyarlılık ve F1-skor değerleri MLP tabanlı modellerden daha düşüktür. Son olarak, KMC tabanlı modeller en düşük doğruluk, kesinlik, hassaslık ve F1-skoru değerlerini üretir. Buna karşın, modellerin duyarlılık değerleri karşılaştırıldığında, en iyiden en kötüye giden makine öğrenmesi sınıflandırıcılarının genel sıralaması MLP, TB, RBF, SVM ve KMC'dir. Ancak modellerin ortalama kesinlik değerleri karşılaştırıldığında, bu metrikte en iyi performansı TB ve SVM'nin gösterdiği, ayrıca TB ve SVM'nin performansının benzer/karşılaştırılabilir olduğu görülmektedir. RBF, MLP ve KMC'nin ortalama kesinlik değerleri karşılaştırıldığında, sınıflandırıcılarının en iyiden en kötüye doğru giden genel sıralamasının RBF, MLP ve KMC olduğu görülmektedir.

Son olarak, COVID-19 tahmini için değişkenlerin öneminde ateş, öksürük, baş ağrısı gibi semptom temelli değişkenler, COVID-19 enfeksiyonunu tahmin etmede en önemli değişkenler olarak bulunmuştur. Öte yandan, yaş, cinsiyet, ırk ve obezite komorbiditesi gibi semptomatik olmayan değişkenlerin modellerin performansı üzerinde ihmal edilebilir bir etkisi vardır.

ACKNOWLEDGMENTS

I wish to express my gratitude to my supervisor Assoc. Prof. Dr. Fatih ABUT for his guidance, patience, advice, encouragements and insight throughout this study.

I would like to thank Prof. Dr. Mehmet Fatih AKAY and Assoc. Prof. Dr. Fatih KILIÇ for their suggestions, comments and serving on my committee.

I would like to thank the Espirito Santo State of Brazil for sharing the dataset with the public used in this thesis.

I would like to thank and acknowledge to Scientific Research Project Unit of Çukurova University for financial support for my thesis (Project Number: FYL-2021-14257).

Last but not the least, I would like to thank my parents for their never-ending support, encouragement and patience throughout my life.

CONTENTS **PAGE**

ABSTRACT I
ÖZ II
EXTENDED ABSTRACT III
GENİŞLETİLMİŞ ÖZET VII
ACKNOWLEDGMENTS XI
CONTENTS……..... XII
LIST OF TABLES……..... XVI
LIST OF FIGURES…… XXIV
LIST OF ABBREVIATIONS XXX
1. INTRODUCTION 1
1.1. Overview of COVID-19 1
1.2. Motivation for COVID-19 Prediction 2
1.3. Types of COVID-19 Prediction Models 3
1.4. Purpose and Contributions of the Thesis 3
1.5. Dataset Generation 4
1.6. Roadmap of the Thesis 6
2. LITERATURE REVIEW 7
2.1. Studies Examining Eurasian Cases 7
2.2. Studies Examining African Cases 16
2.3. Studies Examining American Cases 19
2.4. Studies Examining Intercontinental Cases 27
3. OVERVIEW OF METHODS 35
3.1. Machine Learning Classifiers 35
3.1.1. Multi-Layer Perceptron 35
3.1.2. Tree Boost 37
3.1.3. Radial Basis Function Network 39
3.1.4. Support Vector Machines 40

3.1.5. K-Means Clustering 41
3.2. Feature Selection Algorithms 42
3.2.1. Relief-F 42
3.2.2. Minimum Redundancy Maximum Relevance 43
4. DEVELOPMENT OF PREDICTION MODELS 45
4.1. Methodology 45
4.1.1. Regular Model Creation 45
4.1.2. Feature Selection-Based Model Creation 46
4.1.3. Model Evaluation Metrics 47
4.2. Details of Prediction Models 47
4.2.1. MLP-Based Prediction Models 47
4.2.2. TB-Based Prediction Models 48
4.2.3. RBF-Based Prediction Models 50
4.2.4. SVM-Based Prediction Models 50
4.2.5. KMC-Based Prediction Models 51
5. RESULTS AND DISCUSSION 53
5.1. Results for Models with Unique Combinations of Predictor Variables 53
5.2. Discussions for Models with Unique Combinations of Predictor Variables 57
5.3. Results for Prediction Models Using Double Combinations of the Predictor Variables 60
5.4. Discussion for Prediction Models Using Double Combinations of the Predictor Variables 67
5.5. Results for Models with Triple Combinations of Predictor Variables 69
5.6. Discussions for Models with Triple Combinations of Predictor Variables . 80
5.7. Results for Prediction Models Using Quadruple Combinations of the Predictor Variables 83
5.8. Discussion for Prediction Models Using Quadruple Combinations of the Predictor Variables 96

5.9. Results for Prediction Models Using Quintuple Combinations of the Predictor Variables 99
5.10. Discussion for Prediction Models Using Quintuple Combinations of the Predictor Variables 110
5.11. Results for Prediction Models Using Sextuple Combinations of the Predictor Variables 113
5.12. Discussion for Prediction Models Using Sextuple Combinations of the Predictor Variables 120
5.13. Results for Prediction Models Using Septuple Combinations of the Predictor Variables 123
5.14. Discussion for Prediction Models Using Septuple Combinations of the Predictor Variables 127
5.15. Results for Prediction Models Using Octuple Combinations of the Predictor Variables 129
5.16. Discussion for Prediction Models Using the Octuple Combination of the Predictor Variables 132
5.17. Results for Feature Selection-Based Prediction Models 133
5.17.1. Relief-F Results 133
5.17.2. mRMR Results 139
5.18. Discussion for Feature Selection-based Prediction Models 144
5.19. General Discussion 146
5.20. Comparing Results with Other Studies in the Literature 149
6. CONCLUSION AND FUTURE WORKS 153
REFERENCES 155
CURRICULUM VITAE 161
APPENDIX 163
A.1. Prediction Models Created Using Unique Combinations 165
A.2. Prediction Models Created Using Double Combinations 166
A.3. Prediction Models Created Using Triple Combinations 169

A.4. Prediction Models Created Using Quadruple Combinations 176
A.5. Prediction Models Created Using Quintuple Combinations 184
A.6. Prediction Models Created Using Sextuple Combinations 190
A.7. Prediction Models Created Using Septuple Combinations 194
A.8. Prediction Models Created Using Octuple Combinations 195
A.9. Feature Selection Based Prediction Models 196

LIST OF TABLES **PAGE**

Table 1.1. Descriptive statistics of the dataset 5
Table 2.1. Overview of studies examining Eurasian cases – Part 1 13
Table 2.2. Overview of studies examining Eurasian cases – Part 2 14
Table 2.3. Overview of studies examining Eurasian cases – Part 3 15
Table 2.4. Overview of studies examining African cases. 19
Table 2.5. Overview of studies examining American cases – Part 1 25
Table 2.6. Overview of studies examining American cases – Part 2 26
Table 2.7. Overview of studies examining intercontinental cases – Part 1 33
Table 2.8. Overview of studies examining intercontinental cases – Part 2 34
Table 4.1. List of ranks of predictor variables assigned by mRMR and Relief-F 46
Table 4.2. The intervals for values of the utilized parameters for MLP-based COVID-19 prediction models 48
Table 4.3. The intervals for values of the utilized parameters for TB-based COVID-19 prediction models 49
Table 4.4. The intervals for values of the utilized parameters for RBF-based COVID-19 prediction models 50
Table 4.5. The intervals for values of the utilized parameters for SVM-based COVID-19 prediction models 51
Table 4.6. The intervals for values of the utilized parameters for KMC-based COVID-19 prediction models 51
Table 5.1. Accuracy values for COVID-19 models including the unique combinations of the predictor variables 53
Table 5.2. Precision values for COVID-19 models including the unique combinations of the predictor variables 53
Table 5.3. Recall values for COVID-19 models including the unique combinations of the predictor variables 54

Table 5.4. F1-score values for COVID-19 models including the unique combinations of the predictor variables 54
Table 5.5. Accuracy values for COVID-19 models including the double combinations of the predictor variables – Part 1 60
Table 5.6. Accuracy values for COVID-19 models including the double combinations of the predictor variables – Part 2 61
Table 5.7. Precision values for COVID-19 models including the double combinations of the predictor variables 62
Table 5.8. Recall values for COVID-19 models including the double combinations of the predictor variables 63
Table 5.9. F1-score values for COVID-19 models including the double combinations of the predictor variables 64
Table 5.10. Accuracy values for COVID-19 models including the triple combinations of the predictor variables – Part 1 70
Table 5.11. Accuracy values for COVID-19 models including the triple combinations of the predictor variables – Part 2 71
Table 5.12. Precision values for COVID-19 models including the triple combinations of the predictor variables – Part 1 72
Table 5.13. Precision values for COVID-19 models including the triple combinations of the predictor variables – Part 2 73
Table 5.14. Recall values for COVID-19 models including the triple combinations of the predictor variables – Part 1 74
Table 5.15. Recall values for COVID-19 models including the triple combinations of the predictor variables – Part 2 75
Table 5.16. F1-score values for COVID-19 models including the triple combinations of the predictor variables – Part 1 76
Table 5.17. F1-score values for COVID-19 models including the triple combinations of the predictor variables – Part 2 77

Table 5.18. Accuracy values for COVID-19 models including the quadruple combinations of the predictor variables – Part 1..... 83
Table 5.19. Accuracy values for COVID-19 models including the quadruple combinations of the predictor variables – Part 2..... 84
Table 5.20. Accuracy values for COVID-19 models including the quadruple combinations of the predictor variables – Part 3..... 85
Table 5.21. Precision values for COVID-19 models including the quadruple combinations of the predictor variables – Part 1..... 86
Table 5.22. Precision values for COVID-19 models including the quadruple combinations of the predictor variables – Part 2..... 87
Table 5.23. Precision values for COVID-19 models including the quadruple combinations of the predictor variables – Part 3..... 88
Table 5.24. Recall values for COVID-19 models including the quadruple combinations of the predictor variables – Part 1..... 89
Table 5.25. Recall values for COVID-19 models including the quadruple combinations of the predictor variables – Part 2..... 90
Table 5.26. Recall values for COVID-19 models including the quadruple combinations of the predictor variables – Part 3..... 91
Table 5.27. F1-score values for COVID-19 models including the quadruple combinations of the predictor variables – Part 1..... 92
Table 5.28. F1-score values for COVID-19 models including the quadruple combinations of the predictor variables – Part 2..... 93
Table 5.29. F1-score values for COVID-19 models including the quadruple combinations of the predictor variables – Part 3..... 94
Table 5.30. Accuracy values for COVID-19 models including the quintuple combinations of the predictor variables – Part 1 99
Table 5.31. Accuracy values for COVID-19 models including the quintuple combinations of the predictor variables – Part 2 100

Table 5.32. Accuracy values for COVID-19 models including the quintuple combinations of the predictor variables – Part 3 101
Table 5.33. Precision values for COVID-19 models including the quintuple combinations of the predictor variables – Part 1 102
Table 5.34. Precision values for COVID-19 models including the quintuple combinations of the predictor variables – Part 2 103
Table 5.35. Recall values for COVID-19 models including the quintuple combinations of the predictor variables – Part 1 104
Table 5.36. Recall values for COVID-19 models including the quintuple combinations of the predictor variables – Part 2 105
Table 5.37. F1-score values for COVID-19 models including the quintuple combinations of the predictor variables – Part 1 106
Table 5.38. F1-score values for COVID-19 models including the quintuple combinations of the predictor variables – Part 2 107
Table 5.39. Accuracy values for COVID-19 models including the sextuple combinations of the predictor variables – Part 1 113
Table 5.40. Accuracy values for COVID-19 models including the sextuple combinations of the predictor variables – Part 2 114
Table 5.41. Precision values for COVID-19 models including the sextuple combinations of the predictor variables 115
Table 5.42. Recall values for COVID-19 models including the sextuple combinations of the predictor variables 116
Table 5.43. F1-score values for COVID-19 models including the sextuple combinations of the predictor variables 117
Table 5.44. Accuracy values for COVID-19 models including the septuple combinations of the predictor variables 123
Table 5.45. Precision values for COVID-19 models including the septuple combinations of the predictor variables 123

Table 5.46. Recall values for COVID-19 models including the septuple combinations of the predictor variables 124
Table 5.47. F1-score values for COVID-19 models including the septuple combinations of the predictor variables 124
Table 5.48. Metric values for Model 255 of octuple combination for COVID-19 prediction models. 129
Table 5.49. Accuracy values for Relief-F-based COVID-19 prediction models using different machine learning methods 133
Table 5.50. Precision values for Relief-F-based COVID-19 prediction models using different machine learning methods 134
Table 5.51. Recall values for Relief-F-based COVID-19 prediction models using different machine learning methods 135
Table 5.52. F1-score values for Relief-F-based COVID-19 prediction models using different machine learning methods 136
Table 5.53. Accuracy values for mRMR-based COVID-19 prediction models using different machine learning methods 139
Table 5.54. Precision values for mRMR-based COVID-19 prediction models using different machine learning methods 140
Table 5.55. Recall values for mRMR-based COVID-19 prediction models using different machine learning methods 141
Table 5.56. F1-score values for mRMR-based COVID-19 prediction models using different machine learning methods 142
Table 5.57. Comparison of the results with previous works 151
Table A.5.1. Overview of COVID-19 prediction models along with unique predictor variables 165
Table A.5.2. Overview of COVID-19 prediction models along with double combinations of the predictor variables – Part 1 166
Table A.5.3. Overview of COVID-19 prediction models along with double combinations of the predictor variables – Part 2 167

Table A.5.4. Overview of COVID-19 prediction models along with double combinations of the predictor variables – Part 3 168
Table A.5.5. Overview of COVID-19 prediction models along with triple combinations of the predictor variables – Part 1 169
Table A.5.6. Overview of COVID-19 prediction models along with triple combinations of the predictor variables – Part 2 170
Table A.5.7. Overview of COVID-19 prediction models along with triple combinations of the predictor variables – Part 3 171
Table A.5.8. Overview of COVID-19 prediction models along with triple combinations of the predictor variables – Part 4 172
Table A.5.9. Overview of COVID-19 prediction models along with triple combinations of the predictor variables – Part 5 173
Table A.5.10. Overview of COVID-19 prediction models along with triple combinations of the predictor variables – Part 6 174
Table A.5.11. Overview of COVID-19 prediction models along with triple combinations of the predictor variables – Part 6 175
Table A.5.12. Overview of COVID-19 prediction models along with quadruple combinations of the predictor variables – Part 1..... 176
Table A.5.13. Overview of COVID-19 prediction models along with quadruple combinations of the predictor variables – Part 2..... 177
Table A.5.14. Overview of COVID-19 prediction models along with quadruple combinations of the predictor variables – Part 3..... 178
Table A.5.15. Overview of COVID-19 prediction models along with quadruple combinations of the predictor variables – Part 4..... 179
Table A.5.16. Overview of COVID-19 prediction models along with quadruple combinations of the predictor variables – Part 5..... 180
Table A.5.17. Overview of COVID-19 prediction models along with quadruple combinations of the predictor variables – Part 6..... 181

Table A.5.18. Overview of COVID-19 prediction models along with quadruple combinations of the predictor variables – Part 7..... 182
Table A.5.19. Overview of COVID-19 prediction models along with quadruple combinations of the predictor variables – Part 8..... 183
Table A.5.20. Overview of COVID-19 prediction models along with quintuple combinations of the predictor variables – Part 1 184
Table A.5.21. Overview of COVID-19 prediction models along with quintuple combinations of the predictor variables – Part 2 185
Table A.5.22. Overview of COVID-19 prediction models along with quintuple combinations of the predictor variables – Part 3 186
Table A.5.23. Overview of COVID-19 prediction models along with quintuple combinations of the predictor variables – Part 4 187
Table A.5.24. Overview of COVID-19 prediction models along with quintuple combinations of the predictor variables – Part 5 188
Table A.5.25. Overview of COVID-19 prediction models along with quintuple combinations of the predictor variables – Part 6 189
Table A.5.26. Overview of COVID-19 prediction models along with sextuple combinations of the predictor variables – Part 1 190
Table A.5.27. Overview of COVID-19 prediction along with sextuple combinations of the predictor variables – Part 2 191
Table A.5.28. Overview of COVID-19 prediction along with sextuple combinations of the predictor variables – Part 3 192
Table A.5.29. Overview of COVID-19 prediction along with sextuple combinations of the predictor variables – Part 4 193
Table A.5.30. Overview of COVID-19 prediction models along with septuple combinations of the predictor variables – Part 1 194
Table A.5.31. Overview of COVID-19 prediction models along with septuple combinations of the predictor variables – Part 2 195

Table A.5.32. Overview of COVID-19 prediction models along with octuple combinations of the predictor variables 195
Table A.5.33. Overview of Relief-F-based prediction models for COVID-19 – Part 1 196
Table A.5.34. Overview of Relief-F-based prediction models for COVID-19 – Part 2 197
Table A.5.35. Overview of mRMR-based prediction models for COVID-19 – Part 1 198
Table A.5.36. Overview of mRMR-based prediction models for COVID-19 – Part 2 199

LIST OF FIGURES **PAGE**

Figure 3.1. Pseudo code for TB algorithm 38
Figure 3.2. Algorithm for KMC 41
Figure 3.3. Pseudo code for Relief-F algorithm 42
Figure 5.1. Percentage increase rates in accuracy values of COVID-19 prediction for MLP-based models compared to accuracy values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for unique combinations 55
Figure 5.2. Percentage increase rates in precision values of COVID-19 prediction for MLP-based models compared to precision values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for unique combinations 55
Figure 5.3. Percentage increase rates in recall values of COVID-19 prediction for MLP-based models compared to recall values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for unique combinations 56
Figure 5.4. Percentage increase rates in F1-score values of COVID-19 prediction for MLP-based models compared to F1-score values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for unique combinations 56
Figure 5.5. Percentage increase rates in accuracy values of COVID-19 prediction for MLP-based models compared to accuracy values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for double combinations 65
Figure 5.6. Percentage increase rates in precision values of COVID-19 prediction for MLP-based models compared to precision values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for double combinations 65

Figure 5.7. Percentage increase rates in recall values of COVID-19 prediction for MLP-based models compared to recall values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for double combinations 66
Figure 5.8. Percentage increase rates in F1-score values of COVID-19 prediction for MLP-based models compared to F1-score values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for double combinations 66
Figure 5.9. Percentage increase rates in accuracy values of COVID-19 prediction for MLP-based models compared to accuracy values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for triple combinations..... 78
Figure 5.10. Percentage increase rates in precision values of COVID-19 prediction for MLP-based models compared to precision values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for triple combinations..... 78
Figure 5.11. Percentage increase rates in recall values of COVID-19 prediction for MLP-based models compared to recall values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for triple combinations..... 79
Figure 5.12. Percentage increase rates in F1-score values of COVID-19 prediction for MLP-based models compared to F1-score values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for triple combinations..... 79
Figure 5.13. Percentage increase rates in accuracy values of COVID-19 prediction for MLP-based models compared to accuracy values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for quadruple combinations 94

Figure 5.14. Percentage increase rates in precision values of COVID-19 prediction for MLP-based models compared to precision values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for quadruple combinations 95
Figure 5.15. Percentage increase rates in recall values of COVID-19 prediction for MLP-based models compared to recall values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for quadruple combinations 95
Figure 5.16. Percentage increase rates in F1-score values of COVID-19 prediction for MLP-based models compared to F1-score values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for quadruple combinations 96
Figure 5.17. Percentage increase rates in accuracy values of COVID-19 prediction for MLP-based models compared to accuracy values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for quintuple combinations 108
Figure 5.18. Percentage increase rates in precision values of COVID-19 prediction for MLP-based models compared to precision values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for quintuple combinations 108
Figure 5.19. Percentage increase rates in recall values of COVID-19 prediction for MLP-based models compared to recall values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for quintuple combinations 109
Figure 5.20. Percentage increase rates in F1-score values of COVID-19 prediction for MLP-based models compared to F1-score values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for quintuple combinations 109

Figure 5.21. Percentage increase rates in accuracy values of COVID-19 prediction for MLP-based models compared to accuracy values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for sextuple combinations 118
Figure 5.22. Percentage increase rates in precision values of COVID-19 prediction for MLP-based models compared to precision values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for sextuple combinations 118
Figure 5.23. Percentage increase rates in recall values of COVID-19 prediction for MLP-based models compared to recall values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for sextuple combinations 119
Figure 5.24. Percentage increase rates in F1-score values of COVID-19 prediction for MLP-based models compared to F1-score values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for sextuple combinations 119
Figure 5.25. Percentage increase rates in accuracy values of COVID-19 prediction for MLP-based models compared to accuracy values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for septuple combinations 125
Figure 5.26. Percentage increase rates in precision values of COVID-19 prediction for MLP-based models compared to precision values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for septuple combinations 125
Figure 5.27. Percentage increase rates in recall values of COVID-19 prediction for MLP-based models compared to recall values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for septuple combinations 126

Figure 5.28. Percentage increase rates in F1-score values of COVID-19 prediction for MLP-based models compared to F1-score values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for septuple combinations 126
Figure 5.29. Percentage increase rates in accuracy values of COVID-19 prediction for the MLP-based Model 255 compared to accuracy values obtained by TB-based, RBF-based, SVM-based, and KMC-based Model 255 for the octuple combination..... 130
Figure 5.30. Percentage increase rates in precision values of COVID-19 prediction for the MLP-based Model 255 compared to precision values obtained by TB-based, RBF-based, SVM-based, and KMC-based Model 255 for the octuple combination 130
Figure 5.31. Percentage increase rates in recall values of COVID-19 prediction for the MLP-based Model 255 compared to recall values obtained by TB-based, RBF-based, SVM-based, and KMC-based Model 255 the octuple combination..... 131
Figure 5.32. Percentage increase rates in F1-score values of COVID-19 prediction for the MLP-based Model 255 compared to F1-score values obtained by TB-based, RBF-based, SVM-based, and KMC-based Model 255 for the octuple combination 131
Figure 5.33. Percentage increase rates in accuracies of COVID-19 prediction for MLP-based models compared to accuracies obtained by the rest of the machine learning methods for Relief-F 136
Figure 5.34. Percentage increase rates in precisions of COVID-19 prediction for MLP-based models compared to precisions obtained by the rest of the machine learning methods for Relief-F 137
Figure 5.35. Percentage increase rates in recalls of COVID-19 prediction for MLP-based models compared to recalls obtained by the rest of the machine learning methods for Relief-F 137

Figure 5.36. Percentage increase rates in F1-scores of COVID-19 prediction for MLP-based models compared to F1-scores obtained by the rest of the machine learning methods for Relief-F 138
Figure 5.37. Percentage increase rates in accuracies of COVID-19 prediction for MLP-based models compared to accuracies obtained by the rest of the machine learning methods for mRMR..... 142
Figure 5.38. Percentage increase rates in precisions of COVID-19 prediction for MLP-based models compared to precisions obtained by the rest of the machine learning methods for mRMR..... 143
Figure 5.39. Percentage increase rates in recalls of COVID-19 prediction for MLP-based models compared to recalls obtained by the rest of the machine learning methods for mRMR 143
Figure 5.40. Percentage increase rates in F1-scores of COVID-19 prediction for MLP-based models compared to F1-scores obtained by the rest of the machine learning methods for mRMR..... 144

LIST OF ABBREVIATIONS

COVID-19	:	Coronavirus Disease 2019
GBM	:	Gradient Boosting
H1N1	:	Hemagglutinin Type 1 and Neuraminidase Type 1
KMC	:	K-Means Clustering
MID	:	Mutual Information Difference
ML	:	Machine Learning
MLP	:	Multi-Layer Perceptron
mRMR	:	Minimum Redundancy Maximum Relevance
PCR	:	Polymerase Chain Reaction
RBF	:	Radial Basis Function Network
ReLU	:	Rectified Linear Unit
rRT-PCR	:	Reverse Real-Time Polymerase Chain Reaction
RWBS	:	Repeating Weighted Boosting Search
SARS-CoV-2	:	Severe Acute Respiratory Syndrome Coronavirus 2
SVM	:	Support Vector Machine
TB	:	Tree Boost
WHO	:	World Health Organization

1. INTRODUCTION

1.1. Overview of COVID-19

COVID-19 is an infection that has affected the world since December 31, 2019, and was declared a pandemic by WHO in March 2020. The COVID-19 pandemic has infected 465 million people, claimed more than 6 million lives, and is also known as Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). SARS-CoV-2 causes severe pneumonia and a fatality rate of 2.9%, and most COVID-19 carriers are asymptomatic.

The most common symptoms of COVID-19 are fever, cough, and breathing difficulty, of which 83%, 82%, and 31% of the patients have these symptoms, respectively. However, patients may also have gastrointestinal symptoms such as abdominal pain, diarrhea, and vomiting, of which between 2% and 10% of the patients have these symptoms. COVID-19 mainly harms the lungs and may also harm cardiovascular, central nervous, gastrointestinal, kidney, liver, and ocular systems. SARS-CoV-2's transmission is mainly through the respiratory route like other respiratory viruses. Patients suspected of having SARS-CoV-2 are verified mostly through real-time reverse PCR (rRT-PCR) tests. The disease incubation period varies for each person, but health authorities recommend a 14-day period of active monitoring.

The treatment recommended for the patients varies according to the severity of the cases. For mild cases, it is recommended that patients should stay in bed, but it is also mentioned that blood oxygen saturation monitoring and oxygen therapy with a nasal cannula should be controlled regularly. For moderate cases, it is mentioned that the treatment should include bed rest, supportive treatment to maintain energy supply, maintaining water and electrolyte balance, and monitoring vital signs and oxygen saturations. Still, it is also mentioned that oxygen therapy should be given to patients as needed. They should provide nasal cannula therapy first; if this does not work, treatment with mask oxygen, cannula oxygen therapy,

and inhalation of hydrogen-oxygen is also recommended. It is also mentioned that anti-viral therapy should be given to the patients. Lastly for severe cases, it is mentioned that the main principles of the treatment are active prevention and treatment of complications, prevention of other infections while treating basic diseases, and organ function support in a timely manner (Ciotti et al., 2020).

1.2. Motivation for COVID-19 Prediction

COVID-19 is an infectious disease caused by the coronavirus, also known as SARS CoV-2, a respiratory pathogen first discovered in Wuhan City, China, in December 2019. Since then, COVID-19 has infected more than 465 million people and caused more than 6 million deaths worldwide, which was recognized as a pandemic by WHO on 11 March 2020.

To combat such an infectious and deadly disease, there is a need for models that predict the COVID-19 outbreak as accurately as possible, so governments can make better decisions about the disease. Some of such governmental decisions include (a) arranging a budget to combat disease as effectively as possible, (b) checking the sufficiency of facilities before becoming overfilled, (c) deciding on how many medicines and medical equipment should be produced or imported, and finally (d) how much medical staff is going to be needed.

Studies for developing prediction models for other diseases have already been conducted in the past decades, such as for predicting swine, dengue, H1N1 flu, and influenza. But due to different characteristics, COVID-19's spread is on an unprecedented scale. With the appearance of COVID-19, measures were taken that had almost never been done, such as nationwide curfews, nationwide closing of business and public places, being required to wear a mask outside homes, turning public places like stadiums into temporary hospitals, and many similar measures.

1.3. Types of COVID-19 Prediction Models

Two types of prediction models are currently used to produce COVID-19 outbreak predictions: regression and classification. Regression models predict a continuous quantity output. In contrast, classification is a process of assigning data to appropriate classes, having either two-class (i.e., binary) or multi-class labels to assign data. The main difference between the regression and classification models is that the output variable in regression is numerical (or continuous) while that for classification is categorical. For regression and classification models, the output variable can be predicted in two different ways.

The first way is to find a mapping function to map the input variables x_1, x_2, x_n to the continuous output variable y with f(x) = y. Such types of predictions were referred to as the regular predictions. Alternatively, the output variable can also be predicted based on time series data. Particularly, time series prediction is a process of predicting future values of the data through a sequence of time, based on past values of the data. The principle of the time series prediction method is predicting changes in the future, assuming that values will change similar to how they changed in the past. By combining the regular and time series prediction with the regression and classification models, four types of models were obtained: regular regression models, regular classification models, time series-based regression models, and time series-based classification models.

1.4. Purpose and Contributions of the Thesis

The purpose of this study is to develop various new regular and feature selection-based classification models that effectively predict whether a patient is infected with COVID-19 or not. In this study, a variety of machine learning and feature selection methods were used to create and build a variety of new classification models. Support Vector Machines (SVM), Multi-Layer Perceptron (MLP), Radial Basis Function Network (RBF), K-Means Clustering (KMC), and Tree Boost (TB) Models have been used to build the prediction models. For feature

selection, Relief-F and mRMR algorithms have been preferred. To evaluate the performance of proposed models, well-known performance metrics such as accuracy, precision, recall, and F1-score are used in combination with 10-fold cross-validation.

The contributions and major differences of this thesis from the rest of the studies in the related literature can be summarized as follows:

- This is the first study in the literature that develops new COVID-19 prediction models using machine learning methods combined with feature selection algorithms, with the intention to identify the irredundant and relevant variables in predicting COVID-19 infection. Particularly, two state-of-the-art feature selectors, including Relief-F and mRMR, will be evaluated for the first time for this prediction field.
- By using two feature selectors and five machine learning classifiers on the utilized dataset, this is the most comprehensive study regarding the number of models developed for the prediction of COVID-19 infection.
- By using two feature selectors and five machine learning classifiers, this thesis allows the ranking of these applied methods for prediction of COVID-19 infection.

1.5. Dataset Generation

The dataset used in this study was obtained from the Espirito Santo State Portal of Brazil on May 23, 2021(Data on COVID-19 pandemic, 2021). The dataset contains over 40 columns and has over 1.4 million rows. The dataset contains data that was collected between March 17, 2020, and May 23, 2021, and includes a diverse amount of knowledge such as symptoms of COVID-19 (e.g., fever and cough), biological characteristics of patients (e.g., sex, age, and race), comorbidities of patients (e.g., obesity and diabetes), date-based variables, location-based features (e.g., neighborhood and county), education status, and various test results.

This study aimed to predict whether a patient is infected with COVID-19 by using the patient's symptoms, comorbidities, and biological characteristics. Because of this, the columns that were not planned to use were removed from the dataset. After this step, 17 predictors remained for developing the models. Next, rows with empty values were removed, and string-based values were converted to numeral or binary values. Finally, due to the massive amount of data, a smaller dataset was created by randomly choosing 10.000 positive and 10.000 negative cases.

In the dataset, 0 and 1 values for the "sex" variable represent male and female genders, respectively. Similarly, 0, 1, 2, 3, 4, and 5 values for the "race" variable represent "ignored," "white", "indigenous", "brown", "black", and "yellow", respectively. The descriptive statistics of the dataset are shown in Table 1.1.

Table 1.1. Descriptive statistics of the dataset

Variables	**Minimum**	**Maximum**	**Mean ± Standard Deviation**
Age	0	98	39.24±17.43
Sex	0	1	0.56±0.50
Race	0	5	1.92±1.51
Pregnant	0	1	0.006±0.076
Fever	0	1	0.41±0.49
Breathing difficulty	0	1	0.16±0.37
Cough	0	1	0.54±0.50
Runny nose	0	1	0.39±0.49
Throat pain	0	1	0.33±0.47
Diarrhea	0	1	0.16±0.37
Headache	0	1	0.52±0.50
Lung comorbidity	0	1	0.03±0.18
Cardio comorbidity	0	1	0.13±0.34
Renal comorbidity	0	1	0.005±0.07
Diabetes comorbidity	0	1	0.05±0.21
Smoking comorbidity	0	1	0.02±0.14
Obesity comorbidity	0	1	0.02±0.15

1.6. Roadmap of the Thesis

This thesis contains six chapters. Chapter 2 outlines the previous works in the related literature in detail. Chapter 3 includes basic knowledge about machine learning and feature selection methods used in this thesis. Chapter 4 introduces the methodology along with details of the prediction models developed by utilizing machine learning and feature selection methods. Chapter 5 presents the results and discussions. Finally, Chapter 6 includes the conclusions and future works.

2. LITERATURE REVIEW

In the literature, a variety of studies have been conducted to forecast the COVID-19 infection. This section provides a detailed overview about the related works for predicting the COVID-19 disease. Particularly, this section is divided into four subsections: Studies examining Eurasian cases, studies examining African cases, studies examining American cases, and studies examining intercontinental cases.

2.1. Studies Examining Eurasian Cases

(Aditya Satrio et al., 2021) forecasted COVID-19 spread in Indonesia. To build models, ARIMA (Autoregressive Integrated Moving Average) and PROPHET algorithms were used. The "AR" in ARIMA means autoregressive, "I" means integrated, and "MA" means moving average. ARIMA works by inspecting differences between values instead of using actual values. Usually written in this notation to indicate its parameters: ARIMA (p, d, q), the "p" value means the number of time lags, "d" value means the degree of difference, and "q" means the size of the average window. The dataset also should be checked in order to learn whether the data is stationary or not because ARIMA performs better when data has a stable or consistent pattern over time. In this step usually, ADF (Augmented Dickey-Fuller) method is used. The dataset used in this study was obtained from the Kaggle website and the dataset consists of Serial Number, Observation Date, Province/State, Country/Region, Last Update, Confirmed, Recovered, and Death columns, it had 27618 rows, but after choosing only Indonesian data, only 81 rows remained. Then, they split confirmed, recovery, and death into separate data frames because, ARIMA only uses univariate data. Then they checked the data in order to check whether data is stationary or not because ARIMA performs better when data has a stable or consistent pattern over time. Their finding indicated that data was not stationary, and

after confirmation with ADF, they confirmed that data was not stationary. Then, they used log-scale transformation and time-shifting transformation in order to convert data into stationary. In this study time-series regression models were used. In this study, confirmed, recovered, and death counts were used as variables, and confirmed, recovered, and death counts were predicted as output variables. R^2 (Coefficient of determination), MSE (Mean Square Error), MAE (Mean Absolute Error), and MFE (Mean Forecast Error) metrics have been used as performance metrics. They do a prediction 30-day period for both models from April 22, 2020, to May 21, 2020. It has been concluded that PROPHET generally performed better than ARIMA, but both models become more inaccurate as more days are forecasted.

(Althnian et al., 2020) conducted a study to forecast the susceptibility of individuals based on demographics data, including age, gender, nationality, and location. The dataset was obtained from the Saudi Ministry of Health. The dataset contains data of 229224 patients in Saudi Arabia between 02.03.2020 to 25.04.2020. The predictor variables are the city, confirmatory result, date of birth, gender, nationality, patient's subject id, result date, screening result, and test date. Date of birth was substituted by one of the following age groups: 0-9, 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, and 80+. The city was replaced by 15 different Saudi regions. Because negative test results were overwhelmingly high, random oversampling is used for positive class, and as a result, the resulting database included 50% positive and 50% negative test results. In this study, regular classification models were used, and the variables age, gender, nationality, and location were used as predictors of COVID-10 outbreak. In this study, MLP (Multi-Layer Perceptron), SVM (Support Vector Machine), DT (Decision Tree), and RF (Random Forest) methods were preferred. 5-fold cross-validation was used to train and test the models. Accuracy, precision, recall, F1-score, and AUC (Area under the Curve) metrics were used for performance evaluation. It has been concluded that the DT-based model performs best in terms of all performance metrics. It was also

concluded that susceptibility can be predicted using only the mentioned predictor variables without any medical variables.

(Ceylan, 2020) conducted a study to estimate the prevalence of COVID-19 in Italy, Spain, and France. The dataset was taken from the WHO website and consisted of data collected between 21.02.2020 and 15.04.2020 from Italy, Spain, and France. In this study, 45 samples were used to create a stable and effective time series ARIMA regression model. Various architectures of the ARIMA model are used, and confirmed cases were used as the predictor variable, whereas the number of total cases was predicted. RMSE (Root Mean Square Error), MAE, and MAPE (Mean Absolute Percentage Error) were used as performance metrics. It has been concluded that ARIMA (0, 2, 1), ARIMA (1, 2, 0), and ARIMA (0, 2, 1) are the most suitable models for Italy, Spain, and France, respectively.

(Fang et al., 2020) conducted a study to create a COVID-19 forecasting model to prevent and control COVID-19 outbreaks. The dataset is taken from the John Hopkins University Coronavirus resource center. The dataset contains data between 31.01.2020 and 30.05.2020. In this study, number of confirmed recovered and death cases of Russia were used as variables. Time series regression models were developed based on the ARIMA forecasting method. To verify the fitting effect of the developed ARIMA model, the MAPE metric is preferred. The forecast model for confirmed cases yielded a MAPE value of 0.60, and it has been concluded that the model has high prediction accuracy and is robust. Forecast models for death and recovery yield MAPE values of 3.90 and 2.40, respectively, and it has also been reported that these models were relatively robust.

(Fayyoumi et al., 2020) proposed regular classification models to predict possible patients of COVID-19 using machine learning and statistical models. The utilized dataset was obtained by an online survey at which various eligible participants from Jordan participated, and 105 of those participants were selected for the study. For developing the statistical learning models, LogReg (Logistic Regression) is used, and for building the machine learning models, SVM and MLP

were preferred. The dataset includes various predictor variables, including positive PCR test, age, gender, smoking, positive X-Ray, fever, breathing, diarrhea or vomiting, lack of smell, nasal congestion, dry cough, aches and pain, and sore throat. All models were tested employing 10-fold cross-validation. This study predicts whether the patient is infected with COVID-19 or not. The performance of the models was evaluated using several metrics, including accuracy, sensitivity, specificity, G_Mean (Geometric Mean), and precision. As a result, MLP has been chosen to be the best performing model in terms of the used performance metrics.

(Gupta et al., 2020) proposed a model for predicting the active, death, and cured rate of COVID-19 in India. The dataset contains the active, death, and cured rate for India between 20/01/2020 to 07/05/2020. They used the time series regression model in this study. To build the model, SVM, LinReg (Linear Regression), and prophet forecasting were used. The predictor variables used are active cases, death cases, and cured cases. Active cases, death cases, and cured cases were predicted as the output variables. MAE, MSE, RMSE metrics have been used as performance metrics. It has been concluded that the prophet forecasting model performed better than SVM and LinReg thanks to its features that are not available on the other two models, such as the seasonal component set. Furthermore, it has been reported that SVM becomes the second-best prediction model among those three.

(Önder, 2020) proposed time series regression models that aim to predict the COVID-19 trend. The utilized dataset was collected from the Turkey Ministry of Health as daily confirmed cases. The dataset contains data between March 16, 2020, and March 28, 2020. The models used are Richards Model, Generalized Logistic Growth Model and, the Sub-Exponential growth model. The confirmed case count was used as the predictor variable whereas the reproductive number was predicted as the output variable. Adjusted R^2, RMSE, and AIC (Akaike information criterion) have been used as performance metrics. As a result of this study, sub-exponential

method with a scaling of growth parameter of 0.91 showed the best results among the three.

(Pinter et al., 2020) proposed a COVID-19 prediction model that utilizes hybrid machine learning methods. The dataset includes COVID-19 cases and death rates of Hungary which consists of data collected between March 4, 2020 and April 28 2020, taken from worldometers. For training the model, data collected between March 4, 2020 and April 19, 2020 was utilized, whereas the remaining data was used for validating the predicted results. They used the time series regression model in this study. Total cases and mortality rate were predicted as the target variables. The utilized methods included the Hybrid Multi-Layered Perceptron-Imperialist Competitive Algorithm (MLP-ICA) and ANFIS (Adaptive Network-Based Fuzzy Inference System). Training is done by using two scenarios, the first one (i.e., scenario 1) is trained using odd-numbered days and, the second (i.e., scenario 2) is trained using even-numbered days. The training of ANFIS is done by using three MF (Membership Functions) types: Triangular, Trapezoidal, and Gaussian. RMSE and MAPE were used for comparing these MF results, and it has been concluded that Gaussian MF has the highest performance. The first scenario provided the lowest RMSE results for ANFIS compared to scenario 2 using selected MF because scenario one is more suitable for COVID-19 prediction. Still, it has also been concluded that scenario 2 gives the highest performance for predicting mortality rate using the selected MF type. Also, according to their results, ten neurons MLP architecture yielded the best results for MLP-ICA using both scenarios. But for MLP-ICA, unlike ANFIS, scenario 2 produces the highest accuracy for predicting COVID-19 cases compared to scenario 1. For the mortality rates, neuron number 14 for scenario one and neuron number 18 for scenario 2 yielded the highest accuracy in integrating by ICA method compared to other architectures. Because of it, scenario 1 concluded as the most suitable scenario for mortality rates for MLP-ICA. Evaluations of models were performed by calculating R^2, MAPE, RMSE values. As a result, it has been concluded that different scenarios used for training models do

not create any results that are too different from each other. Also, though both methods showed promising results, MLP-ICA surpassed ANFIS after comparing their results with actual cases.

(Pourghasemi et al., 2020) examined risk factors of COVID-19 and tested the SVM model for mapping areas that have a high risk of human infection in Fars Province, Iran. Also, the growth trend of COVID-19 in Fars Province was analyzed and compared with the growth rate of Iran and various other countries. They developed time series classification models in this study. The area of study is south of Iran, with an area of 122.608 km2. Also, Fars province is split into 36 counties, 93 districts, and 112 cities. The dataset is collected from the Iranian Ministry of Health and Medical Education (IMHME) on April 10, 2020. In this study utilizing data collected from WHO and IMHME, between February 25, 2020, and June 10, 2020, for active cases and from March 2, 2020, to June 10, 2020, for death cases, the growth rate of activity and death cases around the world, Iran, and Fars province were evaluated. The methods included SVM and ARIMA. SVM was used to create a risk map of the COVID-19 outbreak, and ARIMA models were used to examine the patterns of COVID-19 spread in the province. The predictor variables include MTCM (Minimum Temperature of Coldest Month), MTWM (Maximum Temperature of Warmest Month), PWM (Precipitation of Wettest Month), PDM (Precipitation of Driest Month), distance from roads, distance from mosques, distance from hospitals, distance from fuel stations, distance from bus stations, distance from banks, distance from bakeries, distance from attraction sites, distance from ATM's, human footprints, the density of cities, density of villages. For SVM, accuracy and AUC metric were used, and for ARIMA coefficient, standard error, t-statistics, probability metrics were calculated. It has been concluded that climate factors have the least influence on COVID-19 risk mapping and, ATM's, distance from; attraction sites, fuel stations, mosques, and roads, MTCM and, the density of villages and cities have moderate influence, but distances from bus stations, distance from bakeries, and distance from hospitals had the strongest influence. As for

ARIMA, it showed that COVID-19 is not expected to show any explosive process, but it seems infection will keep its general trend for several months.

(Tamhane et al., 2020) published a paper about creating a time series regression model that predicts the case number for the next 20 days. The dataset was obtained from John Hopkins University, consisting of dates between January 1, 2020, and May 25, 2020. To build the models, SVR (Support Vector Regression) and PR (Polynomial Regression) machine learning methods are used. As the predictor variable, the number of cases has been used. First, the data is divided into train and test sets. MAE and MSE are used as performance metrics. This study demonstrated that machine learning methods are very promising and essential tools to deal with the crisis.

Table 2.1. Overview of studies examining Eurasian cases – Part 1

Study	Model	Method	Predictor Variables	Metrics	Target Variable
(Aditya Satrio et al., 2021)	Time Series Regression	Prophet Forecasting Model	Confirmed, recovered, death	R^2, MSE, MAE, MFE	Confirmed, recovered death
(Althnian et al., 2020)	Regular Classification	DT, MLP	Age, gender, nationality, and location	accuracy, precision, recall, F1-score, AUC	Susceptibility
(Ceylan, 2020)	Time Series Regression	ARIMA	Confirmed cases	RMSE, MAE, MAPE	Total cases
(Fang et al., 2020)	Time Series Regression	ARIMA	Confirmed, recovered, death	MAPE	Confirmed, recovered, death

Abbreviations: R^2, Coefficient of determination; MSE, Mean Square Error; MAE, Mean Absolute Error; MFE, Mean Forecast Error; DT, Decision Tree MLP, Multi-Layer Perceptron; AUC, Area under the Curve; ARIMA, Autoregressive Integrated Moving Average; RMSE, Root Mean Square Error; MAPE, Mean Absolute Percentage Error;

Table 2.2. Overview of studies examining Eurasian cases – Part 2

Study	Model	Method	Predictor Variables	Metrics	Target Variable
(Fayyoumi et al., 2020)	Regular Classification	MLP	Positive PCR test, age, gender, smoking, positive X-Ray, fever, breathing, diarrhea or vomiting, lack of smell, nasal congestion, dry cough, aches, and pain, sore throat	Accuracy, sensitivity, specificity, G_Mean, Precision	Positive or negative COVID-19
(Gupta et al., 2020)	Time Series Regression	Prophet Forecasting Model	Confirmed cases, death cases, recovered cases	MAE, MSE, RMSE	Confirmed cases, death cases, recovered cases
(Önder, 2020)	Time Series Regression	sub-exponential growth	Confirmed cases	Adj. R^2, RMSE, AIC	Reproductive number
(Pinter et al., 2020)	Time Series Regression	MLP-ICA	Daily cases, number of deaths	RMSE, MAPE, R^2	Number of cases, mortality rate

Abbreviations: MLP, Multi-Layer Perceptron; G_Mean, Geometric Mean; MAE, Mean Absolute Error; MSE, Mean Square Error; RMSE, Root Mean Square Error; Adj. R^2, Adjusted Coefficient of determination; AIC, Akaike information criterion; MLP-ICA, Multi-Layered Perceptron-Imperialist Competitive Algorithm; MAPE, Mean Absolute Percentage Error; R^2, Coefficient of determination;

Table 2.3. Overview of studies examining Eurasian cases – Part 3

Study	Model	Method	Predictor Variables	Metrics	Target Variable
(Pourghasemi et.al., 2020)	Time Series Classification	SVM, ARIMA	MTCM, MTWM, PWM, PDM, distance from roads, distance from mosques, distance from hospitals, distance from fuel stations, distance from bus stations, distance from banks, distance from bakeries, distance from attraction sites, distance from ATM's, human footprints, the density of cities, the density of villages	Accuracy, AUC, coefficient, standard error, t-statistics, probability	High or low-risk area
(Tamhane et al., 2020)	Time Series Regression	PR, SVR	number of cases	MAE, MSE	Number of cases

Abbreviations: ARIMA, Autoregressive Integrated Moving Average; SVM, Support Vector Machines; AUC, Area under the Curve; MTCM, Minimum Temperature of Coldest Month; MTWM, Maximum Temperature of Warmest Month; PWM, Precipitation of Wettest Month; PDM, Precipitation of Driest Month; PR, Polynomial Regression; SVR, Support Vector Regression. MAE, Mean Absolute Error; MSE, Mean Square Error;

2.2. Studies Examining African Cases

(Gebretensae et al., 2021) proposed a study to predict the future trend of COVID-19 in Ethiopia using ARIMA. The dataset used in this study was obtained from the official website of the Ethiopian Public Health Institute consists of confirmed, recovered, and dead cases in Ethiopia and contains data from March 13, 2020, to August 31, 2020 dates. In this study time-series regression models were used. To build models ARIMA was used. After determining the suitable ARIMA order, they tried to find precise estimates of model parameters utilizing least squares as described by Box and Jenkins. ACF (Autocorrection Function) and PACF (Partial Autocorrection) were used in confirmed cases, in order to check if the data is stationary. ADF tests were used to check whether confirmed and recovered data were stationary, and it has been concluded that data were not stationary. To predict confirmed and recovered, ARIMA (0, 1, 5) and ARIMA (2, 1, 3) were selected respectively based on ACF and PACF graphs. RMSE (Root Mean Square Error), MAPE (Mean Absolute Percentage Error), and BIC (Bayesian information criteria) have been used as performance metrics. It has been concluded that both ARIMA (0, 1, 5) and ARIMA (2, 1, 3) are the best models for confirmed and recovered cases respectively, and confirmed and recovered will increase in Ethiopia in the next 60 days on daily basis.

(Marzouk et al., 2021) proposed a study to forecast the COVID-19 outbreak prevalence in Egypt using deep learning algorithms. The dataset used in this study was obtained from an open-source dataset that was recognized by the Egyptian Ministry of Health and Population, and consists of daily confirmed, recovered, and death cases from Egypt, and consists date between February 14, 2020, and June 30, 2021, and spans over 503 days. The data were split into 90% Training data and 10% testing data. In this study time-series regression models were used. To build models, LSTM (Long Short-Term Memory), CNN (Convolutional neural network), and MLP were used. In this study, confirmed, recovered, and death counts were used as variables, and confirmed, recovered, and death counts were predicted as output

variables. RMSE and R^2 metrics have been used as performance metrics. It has been concluded that LSTM models performed better than CNN and MLP models for 1 week ahead and 1 month ahead predictions because it's able to capture nonlinear patterns in the input data over time and feedback connections characterize the LSTM network in order to propagate the data in the backward pass.

(Ahmed, 2020) published a study examining the prediction of Ethiopian cases to predict the spread of COVID-19. The dataset is taken from the official GitHub repository of the John Hopkins University Center of System Science and Engineering. Only data concerning Ethiopia was used from this dataset. The dataset contains data recorded since 22.01.2020. The dataset was divided into training and testing sets. For training, 3/4 of the dataset was used, whereas the rest was used for testing. The time series regression model was developed based on SVM and PR. The predictor and output variables are confirmed, recovered, and death cases. MAE and MSE metrics were used for performance evaluation. As a result, it has been concluded that SVM performs better than PR in both confirmed, recovered, and death cases.

(Djeddou et al., 2020) predicted new COVID-19 cases in Algeria. The utilized dataset was obtained from the public health database of the Algeria health ministry. They build a time series regression model based on the ELM (Extreme Learning Machine) method. Cumulative confirmed COVID-19 cases, calculated COVID-19 new cases, and index day were used as input variables. The model gives the output of COVID-19 new cases as output. MSE, RMSE, MAE, NSE (Nash–Sutcliffe Coefficient of efficiency), OI (The overall index of model performance), and R^2 metrics were preferred as performance metrics. It has been concluded that the ELM architecture is suitable for this prediction of new COVID-19 cases.

(Saba et al., 2020) created a model for short-term prediction of COVID-19 to help authorities make better decisions. The dataset was obtained from the Egyptian Ministry of Health and Population and consisted of data collected between March 1, 2020, and May 10, 2020. Time series regression models were developed based on

ARIMA and NARANN (Nonlinear Autoregressive Artificial Neural Networks) algorithms. Reported and new cases were used as predictor and target variables, respectively. To performance of the model was evaluated using the MAE, RMSE, R^2, CRM (coefficient of residual mass), and RD (deviation ratio) metrics. They concluded that NARANN performs better than ARIMA for prediction of the new COVID-19 case number.

(Takele, 2020) proposed time series regression model based on the ARIMA method to predict the prevalence of COVID-19 in East African Countries. The motivation of the study is to provide the public health institutions with reliable day-to-day predictions so that proper intervention strategies can be produced and immediate and long-term prevalence trajectory of COVID-19 can be provided. The dataset was taken from the official GitHub repository of John Hopkins University and includes data gathered between March 13, 2020 and June 30, 2020. The countries examined in this study are Ethiopia, Sudan, Djibouti, and Somalia. Confirmed cases and total confirmed cases are used both as predictor and target variables, respectively. AIC, BIC (Bayesian Information Criterion), MAPE, Coefficient, and P-value were used as the model evaluation metrics. ARIMA (1, 2, 1) was chosen as the best model for predicting COVID-19 cases in Ethiopia, ARIMA (2, 1, 1) with drift was chosen as the best model for predicting COVID-19 cases in Djibouti, ARIMA (0, 2, 2) was chosen as the best model for predicting COVID-19 cases in Somalia, and finally ARIMA (2, 2, 1) was chosen as the best model for predicting COVID-19 cases in Sudan.

Table 2.4. Overview of studies examining African cases

Study	Model	Method	Predictor Variables	Metrics	Target Variable
(Gebretensae et al., 2021)	Time Series Regression	ARIMA	Confirmed, recovered	RMSE, MAPE, BIC	Confirmed, recovered
(Marzouk et al., 2021)	Time Series Regression	LSTM	Confirmed, recovered, death	RMSE, R^2	Confirmed, recovered, death
(Ahmed, 2020)	Time Series Regression	SVM	Confirmed, recovered, death	MAE, MSE	Confirmed, recovered, death
(Djeddou et al., 2020)	Time Series Regression	ELM	New COVID-19 cases	MSE, RMSE, MAE, NSE, OI, R^2	New cases
(Saba et al., 2020)	Time Series Regression	NARANN	Reported cases	MAE, RMSE, R^2, CRM, RD	New cases
(Takele, 2020)	Time Series Regression	ARIMA	Confirmed cases	AIC, BIC, MAPE, coefficient, P-value	Total confirmed cases

Abbreviations: ARIMA, Autoregressive Integrated Moving Average; LSTM, Long Short-Term Memory; SVM, Support Vector Machines; RMSE, Root Mean Square Error; MAPE, Mean Absolute Percentage Error; BIC, Bayesian information criterion; R^2, Coefficient of determination; MAE, Mean Absolute Error; MSE, Mean Square Error; ELM, Extreme Learning Machines; NSE, Nash–Sutcliffe Coefficient of efficiency; OI, The overall index of model performance; NARANN, Nonlinear Autoregressive Artificial Neural Networks; CRM, coefficient of residual mass; RD, deviation ratio; AIC, Akaike information criterion

2.3. Studies Examining American Cases

(Luo et al., 2021) proposed a study to predict the future trend of COVID-19 in the US. The dataset used in this study was obtained from the WHO website, and the dataset was in time-series format. Due to the US did not initiate isolation and treatment measures between January and March, new confirmed cases from April 1, 2020, to September 30, 2020, were utilized as modeling objects. The data were split

into 90% Training data and 10% testing data. In this study time-series regression models were used. To build models, LSTM and XGBoost were used. In this study, confirmed cases were used as variables, and confirmed cases for the next 30 days were predicted. MAE, MSE, RMSE, and MAPE metrics have been used as performance metrics. It has been concluded that LSTM performed better than XGBoost as LSTM has a lower MAPE value.

(Santana et al., 2021) conducted a study to assist the detection of COVID-19 based on early symptoms using machine learning. The dataset was created in Brazilian and includes data of 55.676 patients. In this study, regular classification models were developed based on RF, SVM, MLP, KNN (K-Nearest Neighbors), DT, GBM (Gradient Boosting Machine), and XGBoost. Gender, sore throat, dyspnea, fever, cough, headache, taste disorder, olfactory disorder, and health professional were used as the predictor variable. This study predicts whether the patient is infected with COVID-19 or not. Precision, accuracy, recall, and AUC were used as performance metrics. It has been concluded that MLP exhibited the best performance compared to other methods.

(Jojoa et al., 2020) predicted COVID-19 spread in different countries of America. The dataset was downloaded from the open data repository of the European Union. Time series regression-based prediction models using MLP and SVM were used for building the models. This study uses confirmed COVID-19 cases as the predictor and output variable. As for performance measures, CP (Pearson's Correlation Coefficient), MAE, and MPE (Mean Percentage Error) were used. It has been concluded that MLP performs better when an optimization algorithm determines the hyperparameters, but when MLP does not perform well, it is necessary to use SVM instead. MLP showed better performance than SVM for Chile, Mexico, and the USA. However, SVM performs better than MLP for Brazil, Colombia, and Peru using the same performance metrics.

(Moreau, 2020) conducted a study to predict the evolution of the COVID-19 pandemic. The used dataset stems from "Our World in Data Project". The dataset

contains the daily and the total number of confirmed cases and deaths. This dataset includes data starting from February 26, 2020, which is the date of the first confirmed case in Brazil. In this study, the Weibull Distribution model using time series regression is proposed. Daily confirmed cases and daily deaths were used as predictor variables whereas the number of daily new cases and daily new deaths were predicted as output variables. A parameter called diagnostic-death lag is used because it has been previously demonstrated that there is a correlation between lethality rate and diagnostic-death lag. This parameter is obtained by the temporal distance between the peak of daily new cases and daily new deaths. In this study, four different scenarios were conducted. These scenarios were based on the daily number of new deaths. The first, second, third, and fourth scenarios consider 1250, 1500, 1750, and 200 daily new deaths, respectively, at the maximum turning point of the curve. R^2 was preferred as the performance evaluation metric. This study concluded that the first scenario produces the most optimistic results and, the fourth scenario gives the most pessimistic results among those four scenarios. Also, the second and third scenarios yielded results similar to each other and fitted the actual data for daily new cases. Still, because of their similar precision, it's not possible to choose any of these scenarios as the most probable. However, it's mentioned that the diagnostic-death lag value points to a prospective scenario that lies between 1750 and 2000 daily new deaths.

(Silva et al., 2020) conducted a study incorporating the exogenous climatic variables in forecasting models. The datasets for confirmed cases contain cumulative cases in 5 states of the USA and Brazil. The datasets contain data until the date of 28.04.2020. The Brazilian dataset was obtained from an API that collects information about COVID-19 from all Brazilian states. The USA dataset was retrieved from the official repository of Johns Hopkins University whereas the USA climate database was obtained from the National Centers for Environmental Information (NCEI) from the National Oceanic and Atmospheric Administration. The Brazilian climate database was obtained from Instituto Nacional de

Meteorologia. Minimum and maximum temperature and precipitation were used as exogenous climatic inputs for each model. Time series regression models based on BRNN (Bayesian regression neural network), CUBIST (Cubist Regression), KNN, QRF (Quantile Random Forest), SVR, and VMD-based (variational mode decomposition) models were proposed. COVID-19 cases, precipitation, maximum temperature, and minimum temperature were used as predictor variables. The output variable is the number of confirmed cumulative cases. In the study, they first decomposed output variables into five IMF's by performing VMD. They chose lag equal to two by the grid search and used it on the IMF, creating four input lags and exogenous inputs. Also, new data is split into training and test sets. In training, they adopted one-out-cross-validation with a time slice. Second, they trained each IMF with mentioned models, generated five prediction outputs, and named them VMD-BRNN, VMD-CUBIST, VMD-KNN, VMD-QRF, and VMD-SVR. They used a recursive strategy to develop multi-days ahead of COVID-19 case prediction. The model is fitted to do one day ahead prediction; the model uses this forecasting result as input and continues until the desired forecasting horizon is reached. Fourth, they used IP (Improvement Percentage), sMAPE (Symmetric Mean Absolute Percentage Error), and RRMSE (Relative Root Mean Square Error) as metrics. They ranked each of these models for Brazilian states and USA states. For Brazilian states, ranking was VMD-CUBIST, VMD-BRNN, SVR, CUBIST, VMD-SVR, BRNN, VMD-QRF, QRF, VMD-KNN, and KNN. For USA states ranking were VMD-CUBIST, BRNN, CUBIST, SVR, VMD-BRNN, VMD-SVR, VMD-QRF, QRF, KNN, and VMD-KNN. They also mentioned that six-day-ahead hybrid models were more suitable tools than non-decomposed models, and also noted that climatic variables indeed influence the accuracy when predicting COVID-19 cases.

(Souza et al., 2020) conducted a study to make a prognosis or early identification of COVID-19 patients at increased risk of developing severe symptoms. The dataset is obtained from the Espírito Santo state portal. The dataset was made up of 13690 patients who were tested positive for COVID-19. Only closed

cases due to death and recovery were used in this study because the objective of their study was to predict the outcome of the disease. 4826 patient data were used for the training model, and 3617 patient data were used for validation. In this study, regular classification models were developed based on LogReg, Linear Discriminant Analysis, Naïve Bayes, K-Nearest Neighbors, Decision Trees, XGBOOST, and SVM. Fever, respiratory distress, cough, runny nose, sore throat, diarrhea, headache, pulmonary disease, cardiac disease, kidney, diabetes, smoking, obesity, hospitalization and "is patient older than 60" were used as predictor variables. The target variable being predicted indicates whether a patient is died or recovered from COVID-19. ROC (Receiver Operating Characteristic) AUC, PR (Precision-Recall Curve) AUC, precision, recall and, F1-score were used as performance evaluation metrics. Two different experiments were done to evaluate models. As a result of this study, it has been concluded that the best performing model regarding ROC AUC and PR AUC metrics was LogReg, Linear Discriminant Analysis, XGBOOST, and SVM.

(Wollenstein-Betech et al., 2020) proposed regular classification models to predict the following events: Hospitalization, mortality, need for ICU (Intensive Care Unit), and need for a ventilator. The utilized dataset has been provided by the Mexican government to the general public. This dataset contains demographic information such as age, nationality, location, the use of an indigenous language, and pre-existing conditions of patients such as diabetes, chronic obstructive pulmonary disease, immunosuppression, pregnancy, asthma, hypertension, obesity, chronic renal failure, "is using a tobacco" and other prior diseases. The dataset contains data of 91179 patients. 20737 of those patients have positive COVID-19 tests, whereas 15445 of them are waiting for the result of the test. In this study, negative COVID-19 test results were not used for training the models. This study used sparse SVMs, sparse LogReg, Random Forests, and XGBoost models. Accuracy, weighted F1-score, and AUC were used as performance evaluation metrics. It was concluded that according to SVMs and LogReg models, the features that contribute the most to

predicting the hospitalization of the patient were age, gender, chronic renal insufficiency, diabetes, immunosuppression, and pregnancy. The other variables have a smaller impact on the prediction model. The most important features for predicting mortality were age, test status, immunosuppression, and pregnancy. Similarly, the most important features for predicting ICU need were the development of pneumonia (if available), cardiovascular disease, asthma, and test results. Finally, the most important features for predicting the need for a ventilator were: ICU and pneumonia (if available), age, gender, cardiovascular disease, obesity, pregnancy, and the test results.

Table 2.5. Overview of studies examining American cases – Part 1

Study	Model	Method	Predictor Variables	Metrics	Target Variable
(Luo et al., 2021)	Time Series Regression	LSTM	Confirmed cases	MAE, MSE, RMSE, MAPE	Confirmed cases
(Santana et al., 2021)	Regular Classification	MLP	Gender, sore throat, dyspnea, fever, cough, headache, taste disorder, olfactory disorder, health professional	Precision, Accuracy, Recall, AUC	Positive or negative COVID-19
(Jojoa et al., 2020)	Time Series Regression	MLP, SVM	Confirmed cases	CP, MAE, MPE	Confirmed cases
(Moreau, 2020)	Time Series Regression	Weibull distribution	Confirmed new cases, confirmed new deaths	R^2	Daily new cases, daily new deaths
(Silva et al., 2020)	Time Series Regression	BRNN, CUBIST, KNN, QRF, SVR, VMD-based models	COVID-19 cases, precipitation, maximum temperature, minimum temperature	IP, sMAPE, RRMSE	Cumulative confirmed cases

Abbreviations: LSTM, Long Short-Term Memory; MAE, Mean Absolute Error; MSE, Mean Square Error; RMSE, Root Mean Square Error; MAPE, Mean Absolute Percentage Error; MLP, Multi-Layer Perceptron; AUC, Area under Curve; SVM, Support Vector Machines; CP, Pearson's correlation coefficient; MPE, Mean Percentage Error; R^2, Coefficient of determination; BRNN, Bayesian regression neural network; CUBIST, Cubist Regression; KNN, k-nearest neighbors; QRF, Quantile Random Forest; SVR, Support Vector Regression; VMD, Variational Mode Decomposition; IP, Improvement Percentage; sMAPE, Symmetric Mean Absolute Percentage Error; RRMSE, Relative Root Mean Square Error;

Table 2.6. Overview of studies examining American cases – Part 2

Study	Model	Method	Predictor Variables	Metrics	Target Variable
(Souza et al., 2020)	Regular Classification	LogReg, LDA, XGB, SVM	Fever, respiratory distress, cough, runny nose, sore throat, diarrhea, headache, pulmonary disease, cardiac disease, kidney, diabetes, smoking, obesity, hospitalization, is patient older than 60	ROC AUC, PR AUC, precision, recall, F1-score	Is patient recovered or dead
(Wollenstein-Betech et al., 2020)	Regular Classification	SVM, LogReg, random forest, XGB	Age, pregnant, chronic renal insufficiency, diabetes, immunosuppression, chronic obstructive pulmonary disease, obesity, other, hypertension, tobacco use, cardiovascular disease, asthma, gender, ventilator, ICU, test result	Accuracy, weighted F1-score, AUC	Hospitalization, mortality, ICU need, ventilator need

Abbreviations: LogReg, Logistic Regression; LDA, Linear Discriminant Analysis; XGB, XGBOOST; SVM, Support Vector Machines; ROC, Receiver Operating Characteristic; AUC, Area under Curve; PR AUC, Precision-Recall Curve AUC; ICU, Intensive Care Unit

2.4. Studies Examining Intercontinental Cases

(Ayoobi et al., 2021) proposed a study to forecast COVID-19 outbreak and do an in-depth comparison of GRU (Gated Recurrent Unit), LSTM, Conv-LSTM (Convolutional Long Short-Term Memory) with their bidirectional extensions. According to the authors based on their knowledge, Bi-CONV-LSTM and Bi-GRU have not been used as predictors of COVID-19 time series data before. The dataset used in this study was obtained from the WHO website and contained columns such as “Date Reported”, “Country Code”, “Country”, “Who Region”, “Cumulative Cases”, and, “Cumulative Deaths”. The data of Australia and Iran has been used in this study, the data of Australia is between January 25, 2020, to August 19, 2020 dates, and the data of Iran is between January 3, 2020, to October 6, 2020 dates. The data were split into approximately 70% training data and %30 testing data, and about %20 of training data was used for validation. In this study time-series regression models were used. To build models, GRU, LSTM, Conv-LSTM, and bidirectional extensions of these three methods have been used. In this study, “New Cases”, “Cumulative Cases”, “New Deaths”, and “Cumulative Deaths” has been used as predictors, and “new cases” and “new deaths” is predicted in order to predict every one, three, and seven days ahead during next hundred days, and the reason predicting every n days is reducing computational cost while still achieving reasonable amount of performance. MSLE (Mean Squared Log Error), MAPE, RMSLE (Root Mean Squared Log Error), and EV (Explained Variance) metrics has been used as performance metrics. It has been concluded that most of the time bidirectional models outperforms their non-bidirectional counterparts, and for next day prediction for new deaths in both Iran and Australia, Bi-GRU were performed better, for three day ahead predictions of new deaths in Australia, GRU performed better, for three day ahead predictions of new death in Iran, Bi-Conv-LSTM performed better and for seven day ahead predictions for new deaths in Australia, LSTM performed better and seven days ahead for prediction of new deaths in Iran, Bi-Conv-LSTM performed better. For one day ahead predictions of new cases in Australia, LSTM

and Bi-GRU performed better and for three and seven days ahead prediction of new cases in Australia, Conv-LSTM and Bi-Conv-LSTM performed better respectively, as for one and three day ahead new cases predictions of Iran, Bi-GRU performed better and for seven days ahead prediction of new cases in Iran, Bi-Conv-LSTM performed better.

(Bala, 2021) developed a model that predicts COVID-19 outbreaks in any particular country. The dataset is obtained from the "All our World in Data COVID-19 (OWID)" dataset on confirmed cases and confirmed deaths. This dataset is updated daily and published by European CDC under the Oxford University. Time series regression models were developed to predict the total number of cases by using the date and total cases as predictor variables. This database was split into a 70:30 ratio for training and testing, respectively. LinReg, Support Vector Regression, Random Forest Regression, and XGBoost Regression were used to build the models. MAE and RMSE were used as performance evaluation metrics. It has been concluded that the best performing model is based on XGBoost followed by Random Forest Regression, Support Vector Regression, and LinReg.

(Hassan et al., 2021) proposed a study to develop a COVID-19 prediction model that predicts COVID-19 outbreak in order to contribute to humanity getting rid of the COVID-19 pandemic. The dataset used in this study was obtained from the GitHub repository of John Hopkins University and the dataset consists of worldwide daily confirmed, recovered, and death cases data collected between the start of the pandemic and October 4, 2020. The dataset has been divided into two subsets: a training set (220 days) to train models and a testing set (39 days). In this study time-series regression models were used. To build models, MLP, SVM, BN (Bayesian Network), PR, and, LinReg (Linear Regression) were used. In this study, confirmed, recovered, and death counts were used as variables, and confirmed, recovered, and death counts were predicted as output variables. RMSE, MAE, MSE, EV score and R^2 score metrics have been used as performance metrics. It has been

concluded that in general MLP yielded better results in the prediction of both confirmed, recovered, and death cases.

(Yu et al., 2021) proposed a study to create an online artificial intelligence system in order to analyze the dynamic trend of COVID-19, they call this system in short CPAIS (COVID-19 Pandemic AI System). Two datasets were utilized in this study, the first one is from Oxford COVID-19 Government Response Tracker (OxCGRT), and the other one is from John Hopkins University Center for Systems Science and Engineering. Oxford dataset contains government responses to COVID-19 based on a variety of parameters ever since January 1, 2020, contains 20 types of items, their items can be grouped such as containment and closure policies, economic policies, health system policies, and miscellaneous policies. The John Hopkins University dataset contains confirmed, recovered, and death cases containing data for 192 countries ever since January 21, 2020. In this study time-series regression models were used. To build models, ARIMA, FNN (Feedforward Neural Network), MLP, and LSTM were used. In this study, policies used from the OxCGRT dataset with, confirmed, recovered, and death cases from John Hopkins were used as predictors and they used 1-year records (all of the records) from the dataset and used the last 14 days for validation. ME (Mean Error), RMSE, MAE, MPE, and MAPE metrics have been used as performance metrics. It has been concluded that LSTM yielded better results than ARIMA, FNN, and MLP in most of the countries with better accuracy and fewer errors. ARIMA, FNN, and MLP were not stable enough and only yielded competitive results in some specific countries.

(Ardabili et al., 2020) published another study for COVID-19 outbreak prediction using machine learning. They compared several machine learning models using various models for mathematical forecasting to present a comparative analysis of soft computing and machine learning models to predict COVID-19 outbreak. The data consisted of population counts collected from Italy, China, Iran, Germany, and the USA over 30 days. They used the time series regression models. Their comparison of ML optimizer GWO (Gray Wolf Optimizer) outperformed other ML

optimizer methods, with the best accuracy because it had the smallest RMSE and the most significant correlation coefficient. In their comparison of equations, the Logistic model outperformed other models in general in terms of accuracy, which also had the smallest RMSE and largest r-square value. They predict the total number of cases. Training was done using MLP and ANFIS using different neuron numbers for MLP and other MF (Membership Functions) for ANFIS. After comparison, MLP showed the best accuracy.

(Prakash et al., 2020) conducted a study to find which age groups were most affected by COVID-19 using machine learning algorithms. Two datasets were used, and both were obtained from the Kaggle. The first dataset contained data obtained from India, and the second one includes global COVID-19 data. The datasets contain these data; age, state, confirmed Indian national cases, confirmed foreign national cases, cured, deaths, confirmed, date, and month. Regular classification models were developed based on DT, MLR (Multi-LinReg), SVM, XGBoost Classifier, RF Classifier, RF Regressor, KNN+NCA, Gaussian Naïve Bayes (GNB) Classifier, and LogReg. The test dataset ratio is 7:3. The performance evaluation metrics are R^2 and accuracy. It has been reported that 20-30, 30-40, and 40-50 age groups are affected most by COVID-19, and the most successful machine learning methods are the RF Regressor and RF Classifier.

(Punn et al., 2020) predicted the behavior and reachability of COVID-19 by developing time series regression models using various machine learning models. The dataset was retrieved from the official repository of Johns Hopkins University, which includes data collected between January 22, 2020 and April 1, 2020. They used various predictor variables, including province/state, country/region, last update, confirmed, death, and recovered cases. Training and testing are done using real-time data, which uses the number of confirmed, recovered, and death cases as the label for matching date. They predict confirmed, recovered, and death cases. To predict the trend of COVID-19, SVR, PR, DNN (Deep Neural Network), and LTSM with worldwide data is used. The RMSE metric has been preferred to evaluate the

performance of the models. As a result, the PR approach has been reported as the best fit to follow the growing trend.

(Rustam et al., 2020) conducted a study of future prediction of COVID-19 focusing on the number of new cases, deaths, and recoveries to demonstrate the capability of ML models to predict the upcoming patients affected by COVID-19. The dataset was obtained from the GitHub repository provided by the Center for Systems Science and Engineering, John Hopkins University. The dataset contains confirmed cases, deaths, and recoveries in the past number of days the pandemic has started. After the initial data preprocessing step, the dataset has been divided into two subsets: a training set (56 days) to train models and the testing set (10 days). Several time series regression models were built based on LinReg, LASSO (Least Absolute Shrinkage and Selection Operator), SVM, and ES (Exponential Smoothing). To evaluate the performance of the models, R^2, R^2 Adjusted, MAE, MSE, and RMSE metrics were used. Future predictions of new cases, deaths, and recoveries were made and as a result, ES has been reported to outperform all other models in both new cases, deaths, and recoveries. Even after bigger datasets were used, ES still outperformed other methods in accuracy, but other methods exhibited significantly improved performance after using bigger datasets.

(Sahin, 2020) created a COVID-19 forecasting model for predicting daily cases of COVID-19 based on mobility data. The dataset of mobility data was taken from Apple Mobility Trends Reports. The training model is based on data collected between 13.01.2020 and 10.05.2020. The testing models are based on data collected between 13.05.2020 and 20.05.2020. Seven countries were analyzed, including Brazil, France, Germany, Italy, Spain, UK, and the USA. The mobility indices and populations of the countries were used as inputs. Time series regression models were developed based on LinReg, Gaussian Process Regression, SVM, Decision Tree, and Ensemble Learning. MAPE was preferred to evaluate model's performance. It has been concluded that it is impossible to predict daily cases in all countries based on mobility index through machine learning methods.

(Tuli et al., 2020) proposed a real-time, more realistic prediction model that works on a cloud-based computer. The dataset is obtained from "Our World in Data by Hannah Ritchie", which is updated daily from WHO situation reports. For this study, three instance single-core Azure B1's virtual machines with 1 GB RAM, SSD Storage, and 64-bit Windows Server 2016 have been used. To do multiple analysis tasks for predicting various metrics, the HealthFog framework leveraging the FogBus was employed. Robust Weibull Fit and Gaussian Fit were employed to develop time series regression models. The number of cases, new cases, and mortality rate are both input and output variables. MSE, R^2, and MAPE are used as performance evaluation metrics. It has been concluded that the Robust Weibull model works better than the Gaussian Fit model as performance evaluation metrics. It has been concluded that the Robust Weibull model works better than the Gaussian Fit model.

Table 2.7. Overview of studies examining intercontinental cases – Part 1

Study	Model	Method	Predictor Variables	Metrics	Target Variable
(Ayoobi et al., 2021)	Time Series Regression	GRU, Bi-GRU, LSTM, Bi-LSTM, Conv-LSTM, Bi-Conv-LSTM	New cases, cumulative cases, new deaths, cumulative deaths	MSLE, MAPE, RMSLE, EV	New cases, new deaths
(Bala, 2021)	Time Series Regression	XGBoost	Date, total cases	MAE, RMSE	Total cases
(Hassan et al., 2021)	Time Series Regression	MLP	Confirmed, recovered, death	RMSE, MAE, MSE	Confirmed, recovered, death
(Yu et al., 2021)	Time Series Regression	LSTM	Government policies, confirmed, recovered, death	ME, RMSE, MAE, MPE, MAPE	Number of cases
(Ardabili et al., 2020)	Time Series Regression	MLP	Number of cases	RMSE, R	Number of cases
(Prakash et al., 2020)	Regular Classification	RFC, RFR	Age, cured, date, month, confirmed foreign national, confirmed Indian national, deaths, state	R^2, Accuracy	Most affected age group

Abbreviations: GRU, Gated Recurrent Unit; LSTM, Long Short-Term Memory; Conv-LSTM, Convolutional Long Short-Term Memory; ME, Mean Error; MSLE, Mean Squared Log Error; MAPE, Mean Absolute Percentage Error; RMSLE, Root Mean Squared Log Error; EV, Explained Variance; MAE. Mean Absolute Error; RMSE, Root Mean Square Error; MLP, Multi-Layer Perceptron; R, Coefficient of Correlation; RFC, Random Forest Classifier; RFR. Random Forest Regressor; R^2, Coefficient of Determination;

Table 2.8. Overview of studies examining intercontinental cases – Part 2

Study	Model	Method	Predictor Variables	Metrics	Target Variable
(Punn et al., 2020)	Time Series Regression	PR	confirmed, death and recovered cases	RMSE	confirmed cases, recovered cases, death cases
(Rustam et al., 2020)	Time Series Regression	ES	recovery, death, new confirmed cases	MSE, MAE, RMSE, R² Score, R² Adjusted	confirmed cases, recovered cases, death cases
(Sahin, 2020)	Time Series Regression	LinReg, GPR, SVM, DT, EL	mobility index, population	MAPE	daily cases
(Tuli et al., 2020)	Time Series Regression	Robust Weibull Fit	number of cases, mortality rate, new cases	MSE, R^2, MAPE	number of cases, mortality rate, new cases

Abbreviations: PR, Polynomial Regression; RMSE, Root Mean Square Error; ES, Exponential Smoothing; MSE, Mean Square Error; MAE. Mean Absolute Error; R^2, Coefficient of Determination; LinReg, Linear Regression; GPR, Gaussian Process Regression; SVM, Support Vector Machines; DT. Decision Tree; EL, Ensemble Learning; MAPE, Mean Absolute Percentage Error;

3. OVERVIEW OF METHODS

This section contains two subsections. The first section introduces the theoretical details of the ML classifiers. In the second subsection, the working principles of the feature selection algorithms are introduced.

3.1. Machine Learning Classifiers

This section describes the details of the five different ML classifiers, including MLP, TB, RBF, SVM, and KMC.

3.1.1. Multi-Layer Perceptron

MLP is a variant of artificial neural networks and consists of three layers, including an input layer, an output layer, and at least one hidden layer. Every node in layers is called “neurons” and each layer consists of various numbers of neurons. Every neuron is connected with every neuron in the lower and upper layers, and they are not connected with neurons in the same layer. To optimize outputs into desired outputs and reduce the difference between the network output and desired output, every connection in the model has a value named weight. MLP commonly uses a back-propagation algorithm to calculate weights based on descent gradient techniques (Ramchoun et al., 2016).

Also, as an artificial neural network in MLP, every neuron has an activation function. Activation functions are the key elements of artificial neural networks. Rectified Linear Unit (ReLU) and Swish are the most commonly used activation functions. Especially ReLU is the most widely used activation function because of its simplicity and its consistent performance (Misra, 2019).

This study uses the logistic (sigmoid) function for both hidden layer and output layer activation functions. Assume an input layer with n_0 neurons $X = (x_0, x_1, \dots, x_{n_0})$, and a logistic activation function, as shown in (3.1.):

$$f(x) = \frac{1}{1+e^{-x}} \quad (3.1.)$$

To get the network output, the output of every unit in each layer has to be computed. Suppose a set of hidden layers $(h_1, h_2, \ldots\ h_N)$. Suppose that n_i is the neurons number by every hidden layer h_i. The output of the first hidden layer can be formulated as:

$$h_i^j = f(\sum_{k=1}^{n_{i-1}} w_{k,j}^0\ x_k) \quad j = 1, \ldots, n_i \quad (3.2.)$$

The outputs h_i^j of neurons in the hidden layers are calculated as:

$$h_i^j = f\left(\sum_{k=1}^{n_{i-1}} w_{k,j}^{i-1}\ h_{i-1}^k\right)\ i = 2, \ldots, N\ and\ j = 1, \ldots, n_i \quad (3.3.)$$

where $w_{k,j}^i$ is the weight between the neuron k and neuron j in hidden layer i and hidden layer +1, respectively. n_i is the number of the neurons in the i^{th} hidden layer. The output of the i^{th} hidden layer can be formulated as:

$$h_i = \ ^t(h_i^1, h_i^2, \ldots, h_i^{n_i}) \quad (3.4.)$$

The network output is calculated by (3.5.):

$$y_i = f\left(\sum_{k=1}^{n_N} w_{k,j}^N\ \ h_N^k\right) \quad Y = \left(y_1, \ldots, y_j, \ldots, y_{N+1}\right) = F(W, X) \quad (3.5.)$$

where $w_{k,j}^N$ is the weight between neuron k and neuron j in the N^{th} hidden layer and output layer, respectively, n_N is the number of neurons in the N^{th} hidden layer, Y is the vector of the hidden layer, F is the transfer function, and W is the matrix of the weights, as shown in (3.6.):

$$W = \left[W^0, \dots, W^j, \dots, W^N\right] \tag{3.6.}$$

$$W^i = \left(w^i_{j,k}\right)_{\substack{0 \le i \le N \\ 1 \le j \le n_{i+1} \\ 1 \le k \le n_i}} \quad where\ w^i_{j,k} \in \mathbb{R} \tag{3.7.}$$

To simplify, assume all hidden layers in the network are like $n = n_i\ \forall i = 1, \dots, N$. In this section, X represents the input of the neural network, f represents the activation function, W^i represents the matrix of the weights between the i^{th} hidden layer, and $(i+1)^{th}$ hidden layer for $i = 1, \dots, N-1$, W^0 represents the matrix of weights between the input layer and the first hidden layer, and W^N represents the matrix of weights between the Nth hidden layer and output layer (Ramchoun et al., 2016).

3.1.2. Tree Boost

TB is a special type of Gradient Boosting (GB). These models utilize numerous weak learners to create a strong ensemble. Gradient boosting is a repetitive process. It begins its early approximation of a function using a constant offset that fits the data poorly. The fit is improved in each following iteration by fitting a base learner to the negative gradient of a pre-specified loss function. The approximation of the prediction function is updated with the approximation of the negative gradient, and the functions get closer to the best form of the relationships with each following iteration. GB has this definition:

$$X \coloneqq X_1, X_2, \dots, X_p \tag{3.8.}$$

Let the realizations denoted be as $(X_1, Y_1), \dots, (X_n, Y_n)$ and n as the number of sample sites. X is defined as $X \coloneqq (X_1, \dots, X_n)$ and Y is defined as $Y \coloneqq (Y_1, \dots, Y_n)$. GB approximates the optimal prediction function as

$$f^* \coloneqq argmin_f E_{Y,X}[\rho(Y, f(X))] \quad (3.9.)$$

The function of *f** is minimizing the expectation of a loss function ρ. Among all possible prediction functions, f takes $X \coloneqq X_1, X_2, \dots, X_\rho$ as input. The accurate distributions of *X* and *Y* that are needed to derive (3.8.) are usually unknown, so instead, GB minimizes the empirical risk over *f* in (3.9.).

$$R \coloneqq \frac{1}{n}\sum_{i=1}^{n} p(Y_i, f(X_i)) \quad (3.10)$$

X and Y sample points are represented by $(X_1, \dots, X_n)$ and $(Y_1, \dots, Y_n)$, respectively. *R* represents an estimation of theoretical risk (Maloney et al., 2012).

TB, also called Stochastic Gradient Boosting, randomly draws a subsample of training data from the entire training dataset without replacements. Furthermore, TB is used for fitting the base learners rather than a full sample and calculates the model update for the current iteration (Friedman, 1999). Figure 3.1 gives pseudo code for TB algorithm.

1. $F_0(X) = argmin_\gamma \sum_{i=1}^{N} \psi(y_i, \gamma)$
2. for m = 1 to M do:
3. $\{\pi(i)\}_1^N = rand_perm\ \{i\}_1^N$
4. $\tilde{y}_{\pi(i)m} = -\left[\frac{\partial \psi(y_{\pi(i)}, F(X_{\pi(i)}))}{\partial F(X_{\pi(i)})}\right]_{F(X)=F_{m-1}(x)}, i = 1, \tilde{N}$
5. $\{R_{lm}{}_1^L = L - terminal\ node\ tree(\{\tilde{y}_{\pi(i)m}, X_{\pi(i)}\}_1^{\tilde{N}})$
6. $\gamma lm = argmin_\gamma \sum_{x_{\pi(i)} \in R_{lm}} \psi(y_{\pi(i)}, F_{m-1}(X_{\pi(i)}) + \gamma)$
7. $F_m(x) = F_{m-1}(x) + \nu \cdot \gamma_{lm} 1(x \in R_{lm})$
8. endfor

Figure 3.1. Pseudo code for TB algorithm (Friedman, 1999)

3.1.3. Radial Basis Function Network

RBF is a type of neural network that consists of three layers of neurons. The basis of the RBF originated from the theory of function approximation. The Euclidean distance is calculated out of the point being assessed to the center of each neuron, and a radial basis function is applied to the distance to calculate the weight for each neuron.

The basis of the RBF is the radius distance. Its output hinges on the distance of the input from a given kept vector. That signifies in case the distance from the input vector $\vec{x}$ to the center $\vec{\mu}_i$ of each RBF φ_j, that means $\|\vec{x} - \vec{\mu}_j\|$ is equal to 0, then the contribution of this point is 1, whereas, in the case of $\|\vec{x} - \vec{\mu}_j\|$ increase, the contribution leans to 0.

RBFs are mostly made of three layers: the input layer, hidden layer, and output layer. Input layers can have more than a single predictor variable and, each neuron is associated with one independent neuron and feeds values to the hidden layer neurons. The hidden layer neurons can have multiple numbers of neurons. Every neuron in the hidden layer is made of an RBF centered at a point, relying on the dimensionality of the input and output predictor variables. The hidden layer activation functions are obtained by the basis functions $\varphi_j(\vec{x}, \vec{\mu}_j, \sigma_j)$, which hinges on the parameters $\{\vec{\mu}_j \sigma_j\}$ and input activations $\{\vec{x}\}$ in a non-standard manner.

$$\varphi_j(\vec{x}) = exp\left(-\frac{\|\vec{x}-\vec{\mu}_j\|^2}{2\sigma_j^2}\right) \tag{3.11.}$$

The radius of the RBF function may vary for each dimension. The training process of the network decides the centers and the spreads. As for the output layer, each neuron is associated with weights that can be written as $w_1, \ldots, w_n$. The output of hidden layers is multiplied by the weights associated with the neurons and transferred to the output layer, which adds up the weighted values and gives the

resulting sum as the network's output. Finally, a bias value is multiplied by a weight and fed to the output layer (Dash et al. 2016).

3.1.4. Support Vector Machines

SVM is a robust algorithm with strong regularization properties. SVM minimizes empirical classification error and maximizes the geometric margin at the same time. Because of this, SVMs can also be called maximum margin classifiers. The basis of SVM is structural risk minimization. SVM maps the input vector to a higher-dimensional space in which a maximal separating hyperplane is built. Two hyperplanes are built on both sides of the hyperplane that separate data. The separating hyperplane maximizes the distance between two parallel hyperplanes. It can be presumed that the larger margin or distance between these parallel hyperplanes, the better generalization error of the classifier will be. Assume data points as $\{(x_1, y_1), \ldots (x_n, y_n)\}$ and y_n= 1/-1 is a constant denoting the class that x_n belongs, and n equals to the number of samples. Every x_n is a p-dimensional real vector. Scaling is essential to guard against variables with larger variance. To examine training data by dividing hyperplane, which takes $w \cdot x + b = 0$. In this equation, b represents the scalar, w represents the p-dimensional vector, and w points perpendicular to the separating hyperplane. The margin can be increased by adding the offset parameter b. Without b, the hyperplane has to pass through the origin. Because of the interest in increasing the maximum margin, parallel hyperplanes are essential. Parallel hyperplanes can be formulated as:

$$w \cdot x + b = 1 \qquad (3.12.)$$
$$w \cdot x + b = -1$$

In case training data can be separated linearly, these hyperplanes can be selected thus that there are no points between them, and their distance can be tried to be

maximized. The distance between the hyperplanes can be formulated as: $2/|w|$. Thus, $|w|$ must be minimized. Thus, to excite the data points:

$$w \cdot x_i - b \geq 1 \text{ or } w \cdot x_i - b \leq -1 \quad (3.13.)$$

It can also be formulated as such:

$$y_i(w \cdot x_i - b) \geq 1\,, 1 \leq i \leq n \quad (3.14.)$$

This study uses non-linear SVM; thus, a kernel function must be used. In this study, RBF kernel is utilized. RBF kernel can be formulated as such (Bhavsar et al. 2012):

$$K(x_i, x_j) = \exp\left(-\gamma \left\| x_i - x_j \right\|^2\right), \gamma > 0 \quad (3.15.)$$

3.1.5. K-Means Clustering

KMC is an iterative machine learning method usually used in fields such as information retrieval, pattern recognition, and computer vision. KMC selects n data points into k predetermined clusters to group similar data points. KMC is an iterative ML method that sets every data point recalculated by taking its average (Shukla et al., 2014). The algorithm for KMC is given in Figure 3.2.

1. A starting clustering created by choosing k random cluster centers from the dataset.
2. For every data point, compute the distance from all cluster centers and set it to closest cluster center.
3. Recompute the new cluster centers by the average of all data points assigned to the clusters.
4. Repeat 2 until the convergence

Figure 3.2. Algorithm for KMC (Shukla et al. 2014)

3.2. Feature Selection Algorithms

This subsection briefly introduces the theoretical foundations of the Relief-F and mRMR feature selectors.

3.2.1. Relief-F

The Relief-F algorithm is a feature selection algorithm based on instances. It randomly chooses an instance and searches for the nearest neighbors of the same class. Unlike the original Relief algorithm, the Relief-F algorithm can deal with incomplete data and is not restricted to two-class problems.

The pseudo-code of the Relief-F algorithm is given in Figure 3.3. Like the original Relief algorithm, the Relief-F algorithm also sets weights to 0 (first line). Then also, like the Relief algorithm, an instance (R_i) is randomly chosen (third line); however, unlike the Relief algorithm, the Relief-F algorithm looks for k nearest neighbors of the same class, which is called H_j (fourth line). Also, it looks for nearest misses M_j (C) which are found by looking for k nearest neighbors of different classes (Fifth and sixth line). Then W[A] updates for all attributes A according to their values for R_i and H_j (Robnik-Šikonja and Kononenko, 2003).

Input: for each training instance, a vector of attribute values and the class value
Output: The vector W of estimations of the qualities of attributes

1. Set all weights W[A] := 0.0;
2. **for i** :=1 **to** m **do begin**
3. randomly select an instance R_i;
4. find k nearest hits H_j;
5. **for** each class C ≠ class(R_i) **do**
6. from class C find k nearest misses M_j(C);
7. **for** A := 1 **to** a **do**
8. $W[A] := W[A] - \sum_{j=1}^{k} diff(A, R_i, H_j)/m.k +$
9. $\sum_{C \neq class(R_i)} [\frac{P(C)}{1-P(class(R_i))} \sum_{j=1}^{k} diff(A, R_i, M_j(C))]/(m.k.)$;

end;

Figure 3.3. Pseudo code for Relief-F algorithm (Robnik-Šikonja and Kononenko, 2003)

3.2.2. Minimum Redundancy Maximum Relevance

There are two main aims in the methodology of mRMR: first is not using variables that are highly correlated among each other, i.e., redundancy between features should be taken into account, so keeping features that are maximally dissimilar to each other (Acid et al., 2011). The formula shown in (3.16) could be used to calculate redundancy among the variables.

$$W_I(S) = \frac{1}{|S|^2} \sum_{X_i, X_j \in S} MI(X_i, X_j) \quad (3.16.)$$

where $MI(X_i, X_j)$ provides the measure of mutual information between variables X_i and X_j.

The second aim is that minimum redundancy should be augmented by utilizing the maximum relevance criterion of the features concerning the class variable. The global relevance of the variables in *S* regarding C could be calculated as:

$$V_I(S) = \frac{1}{|S|} \sum_{X_i \in S} MI(C, X_i) \quad (3.17.)$$

The easiest way to get a good subset of features by combining redundancy and relevance could be given as follows:

$$S^* argmax_{S \subseteq U} (V_I(S) - W_I(S)) \quad (3.18.)$$

It is meaningless to look for the best subset by an exhaustive search. The chosen subset is acquired in an incremental way, beginning with the feature having a maximum value of $MI(C; X_i)$ $(S_0 = \{X_{i_0}\})$ and gradually adding to the current subset S_{m-1}, which is the feature it maximizes (Acid et al., 2011).

$$\left(\underset{X_j \in U \setminus S_{m-1}}{max} \left(MI(C, X_j) - \frac{1}{m-1} \sum_{X_i \in S_{m-1}} MI(X_j, X_i) \right) \right) \quad (3.19)$$

4. DEVELOPMENT OF PREDICTION MODELS

This section contains two subsections. The first subsection explains the methodology of regular and feature selection-based models and model evaluation metrics. The second subsection introduces the preferred parameters of prediction models.

4.1. Methodology

This section contains detailed information about two types of model creation used in this study. First, the creation of regular models has been explained, and then, the feature selection-based models have been explained. Both regular models and feature selection-based models

All the models in this study were trained using the DTREG (Sherrod, 2016) software. DTREG is a commercial predictive modeling software that accommodates a variety of ML techniques that allows the development of both classification and regression models.

4.1.1. Regular Model Creation

The common variables appearing in all regular models have been chosen as age, sex, fever, breathing difficulty, cough, runny nose, throat pain, diarrhea, and headache. By using the common variables and the unique, double, triple, quadruple, quintuple, sextuple, septuple, and octuple combinations of the predictor variables pregnancy, race, lung comorbidity, renal comorbidity, cardio comorbidity, diabetes comorbidity, smoking comorbidity and obesity comorbidity, variety of new regular prediction models have been created. The combinations of these predictor variables resulted in 8, 28, 56, 70, 56, 28, 8, and 1 different models for unique, double, triple, quadruple, quintuple, sextuple, septuple, and octuple combinations of predictor variables respectively. MLP, TB, RBF, SVM, and KMC have been used with each combination. The performance of the models was evaluated using accuracy,

precision, recall, and F1-score metrics. The generalization errors have been assessed by 10-fold cross-validation. Due to the limited amount of space, all regular prediction models and their predictor variables have been included in the Appendix (in Table A.5.1. through Table A.5.32).

4.1.2. Feature Selection-Based Model Creation

By utilizing the mRMR and Relief-F feature selectors, the rank of each variable has been calculated and sorted in descending order by their ranks, as shown in Table 4.1. Then, each variable with the lowest rank was removed successively. By doing so, 17 mRMR-based and 17 Relief-F-based prediction models were created, as included in the Appendix (in Table A.5.33. through Table A.5.36).

Table 4.1. List of ranks of predictor variables assigned by mRMR and Relief-F

Rank	Predictor Variables Ranked by mRMR	Predictor Variables Ranked by Relief-F
1	Fever	Age
2	Cough	Race
3	Headache	Fever
4	Age	Cough
5	Breathing difficulty	Headache
6	Obesity comorbidity	Cardio comorbidity
7	Throat pain	Runny nose
8	Diarrhea	Diarrhea
9	Runny nose	Throat pain
10	Pregnant	Breathing difficulty
11	Smoking comorbidity	Sex
12	Renal comorbidity	Diabetes comorbidity
13	Race	Obesity comorbidity
14	Lung comorbidity	Pregnant
15	Sex	Lung comorbidity
16	Diabetes comorbidity	Smoking comorbidity
17	Cardio comorbidity	Renal comorbidity

4.1.3. Model Evaluation Metrics

Prediction errors have been calculated using accuracy, precision, recall, and F1-score metrics, which are given in (4.1.), (4.2.), (4.3), and (4.4), respectively.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4.1.)$$

$$Precision = \frac{TP}{TP+FP} \quad (4.2.)$$

$$Recall = \frac{TP}{TP+FN} \quad (4.3.)$$

$$F1 - \text{Score} = 2 \times \frac{PrecisionxRecall}{Precision+Recall} \quad (4.4.)$$

TP, TN, FP, and FN refer to true positive, true negative, false positive, and false negative, respectively. 10-fold cross-validation has been used to assess the generalization errors in all models.

4.2. Details of Prediction Models

This section explains the preferred parameter values of five different ML classifiers.

4.2.1. MLP-Based Prediction Models

Optimizing the MLP-based models required four different parameters, which are given in Table 4.2. These parameters are the number of the hidden layers, hidden layer activation functions, output layer activation functions, and the number of neurons in the hidden layer.

Table 4.2. The intervals for values of the utilized parameters for MLP-based COVID-19 prediction models

Parameter	Range
Number of the hidden layers	1
Hidden layer activation function	Logistic
Output layer activation function	Logistic
Number of neurons in the hidden layer	[2-20]

The number of the hidden layers in this study is only one. The necessary number of hidden layers may vary according to the study used, but usually, one layer is enough for most kinds of data.

The logistic function is chosen for hidden and output layer activation functions. Activation functions compute the weighted sum and add bias to it to create non-linear outputs for the neurons. The logistic function, also known as the sigmoid function, given in (4.1.), produces "S" shaped graphs.

$$f(x) = \frac{1}{1+e^{-x}} \tag{4.1.}$$

The DTREG program optimizes the number of neurons in the hidden layer. The program has chosen values between 2 to 20. Neurons use the activation functions to produce output, so its number is important for the model (Sherrod, 2016).

4.2.2. TB-Based Prediction Models

Optimizing the TB-based models required five parameters, which are given in Table 4.3. These parameters are the number of trees in the series, the depth of individual trees, the proportion of rows for each tree, the influence trimming factor, and the minimum size node to split.

Table 4.3. The intervals for values of the utilized parameters for TB-based COVID-19 prediction models

Parameter	Range
Number of trees in the series	[10-400]
Depth of individual trees	5
Proportion of rows for each tree	0.5
Influence trimming factor	0.01
Minimum size node to split	10

The DTREG program optimizes the number of trees in the series and chooses values between 10 to 400 as the number of trees in the series. Usually, a larger number of trees is better, but it will increase the training time and stop improving at some point.

The depth of the individual trees is chosen as 5. The depth of the individual tree represents the depth of each tree in the model. The depth is also crucial because the number of nodes depends on the depth of the trees. The number of nodes in the terminal equals 2^n.

The proportion of rows for each tree represents the proportion of random subsample to train each individual tree. In this study, 0.5 was chosen as the proportion of rows for each tree.

The influence trimming factor is used for optimizing the speed of the models and does not affect model performance visibly. Influence trimming allows excluding the next tree build process rows with low residual values. In this study, a 0.01 influence trimming factor is used, so this parameter setting ignores rows whose total residual represents only 1% of the total residual weight.

The minimum size node to split represents the smallest number of nodes required to split the trees. In this study, the minimum size to split is chosen as 10 (Sherrod, 2016).

4.2.3. RBF-Based Prediction Models

Optimizing the RBF-based models required four parameters, which are given in Table 4.4. These parameters are the maximum number of neurons, radius, lambda, and population size.

As for all neural networks, neurons are the part of the models that use activation functions, weighs, and bias to give output. In this study, the maximum number of neurons is set to 100.

The radius parameter represents the spread of the neurons. The DTREG program chooses values between 0.01 to 400 as the radius parameter.

The lambda parameter is used to calculate weights as neurons are added to the network. It is a regularization parameter optimized by the DTREG program, and its value is between 0.001 to 10.

Population size is a parameter used by the Repeating Weighted Boosting Search algorithm. A population of candidate neurons is created with random centers and spreads during the initial phase of the search. The population size determines the number of candidate neurons (Sherrod, 2016).

Table 4.4. The intervals for values of the utilized parameters for RBF-based COVID-19 prediction models

Parameter	Range
Maximum number of neurons	100
Radius	[0.01-400]
Lambda	[0.001-10]
Population size	200

4.2.4. SVM-Based Prediction Models

Optimizing the SVM-based models required two parameters, which are given in Table 4.5. These parameters are cost and gamma.

Table 4.5. The intervals for values of the utilized parameters for SVM-based COVID-19 prediction models

Parameter	Range
Cost (C)	[0.1-2000]
Gamma (γ)	[0.001-20]

Both cost and gamma parameters are calculated by grid search. In this study, the RBF kernel is used for SVM. The cost parameter is used for optimizing the margin used in SVM. Decides how soft or hard the margin is. Large C values may cause overfitting, or small C may cause underfitting. C is a regularization parameter. Gamma parameter is used for the optimization of the models in non-linear SVM. Gamma represents the influence of a single training example. High or low gamma may reduce the performance of the models. Large gamma values may cause overfitting in the models (Sherrod, 2016).

4.2.5. KMC-Based Prediction Models

Optimizing the KMC-based models required two parameters, which are given in Table 4.6. These parameters are the number of clusters and the search step. KMC is a method of separating the data into k number of clusters automatically. In this study number of clusters and search step is used as parameters.

The number of clusters (K) represents how many clusters will be used in the model. It is the primary parameter used by the KMC. This parameter is optimized by the DTREG program.

The search step parameter represents the number of clusters between each step (Sherrod, 2016).

Table 4.6. The intervals for values of the utilized parameters for KMC-based COVID-19 prediction models

Parameter	Range
Number of clusters	[2-200]
Search step	1

5. RESULTS AND DISCUSSION

This section presents the results and discussion of regular and feature selection-based prediction models.

5.1. Results for Models with Unique Combinations of Predictor Variables

Table 5.1 through Table 5.4 show the accuracy, precision, recall, and F1-score values for COVID-19 prediction models including the unique combinations of the predictor variables, respectively. Figure 5.1 through 5.4 illustrates the percentage increase rates in accuracy, precision, recall, and F1-score values for COVID-19 prediction models.

Table 5.1. Accuracy values for COVID-19 models including the unique combinations of the predictor variables

Models	Accuracy (%)				
	MLP	TB	RBF	SVM	KMC
Model 1	70.71	69.92	69.75	70.50	66.47
Model 2	70.41	69.81	69.83	69.83	65.86
Model 3	70.59	69.84	69.67	70.16	66.42
Model 4	70.65	69.80	69.58	70.07	66.15
Model 5	70.73	70.05	69.74	70.38	66.45
Model 6	70.88	69.99	69.62	70.42	66.29
Model 7	70.63	69.92	69.76	70.34	66.02
Model 8	70.55	69.83	69.77	70.32	65.86

Table 5.2. Precision values for COVID-19 models including the unique combinations of the predictor variables

Models	Precision (%)				
	MLP	TB	RBF	SVM	KMC
Model 1	65.74	66.14	66.10	65.42	65.03
Model 2	65.24	66.26	66.29	66.48	64.49
Model 3	66.06	65.89	66.13	65.01	65.25
Model 4	66.05	66.05	66.02	64.87	64.69
Model 5	65.84	66.23	66.11	65.30	65.04
Model 6	65.77	66.19	66.07	65.19	65.08
Model 7	66.28	66.15	66.18	65.25	65.19
Model 8	65.79	66.07	66.10	65.10	64.95

Table 5.3. Recall values for COVID-19 models including the unique combinations of the predictor variables

Models	Recall (%)				
	MLP	TB	RBF	SVM	KMC
Model 1	86.51	81.62	81.06	87.01	71.30
Model 2	87.37	80.71	80.68	79.98	70.62
Model 3	84.69	82.26	80.65	87.30	70.25
Model 4	84.98	81.45	80.73	87.57	71.13
Model 5	86.18	81.82	81.02	87.01	71.14
Model 6	87.04	81.72	80.64	87.62	70.30
Model 7	83.99	81.61	80.84	87.05	68.75
Model 8	85.62	81.51	81.15	87.61	68.92

Table 5.4. F1-score values for COVID-19 models including the unique combinations of the predictor variables

Models	F1-score (%)				
	MLP	TB	RBF	SVM	KMC
Model 1	74.71	73.07	72.82	74.68	68.02
Model 2	74.70	72.78	72.78	72.61	67.41
Model 3	74.22	73.17	72.67	74.53	67.66
Model 4	74.33	72.95	72.63	74.53	67.76
Model 5	74.65	73.20	72.81	74.61	67.95
Model 6	74.93	73.14	72.63	74.76	67.59
Model 7	74.09	73.07	72.78	74.59	66.92
Model 8	74.41	72.98	72.86	74.70	66.87

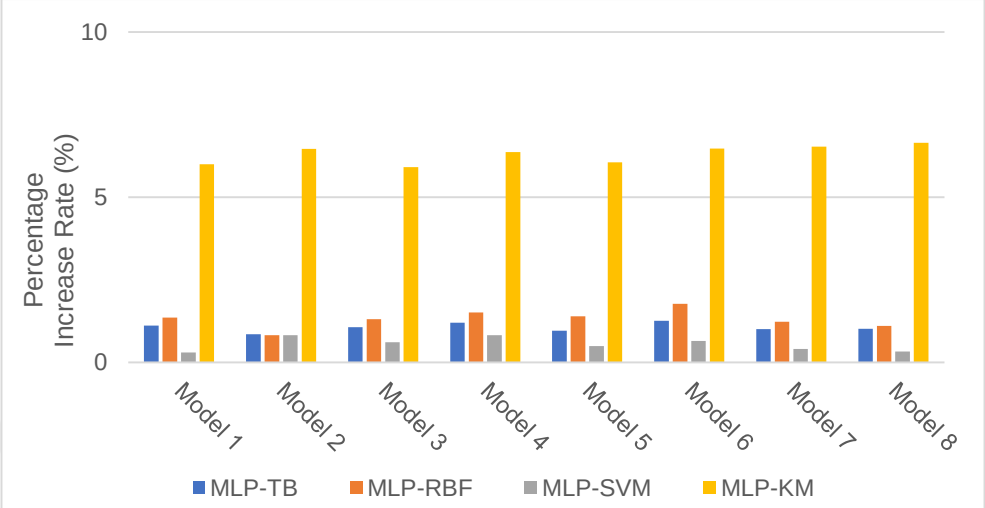


Figure 5.1. Percentage increase rates in accuracy values of COVID-19 prediction for MLP-based models compared to accuracy values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for unique combinations

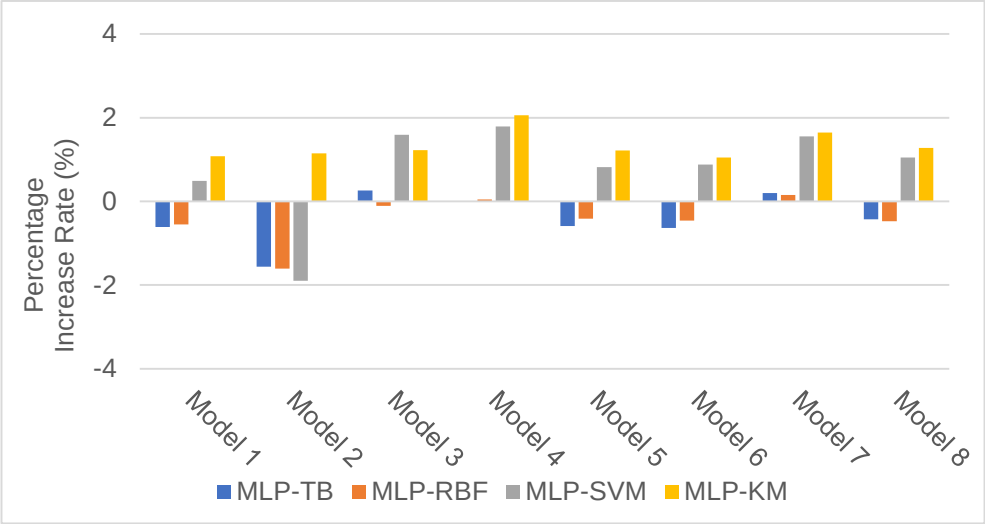


Figure 5.2. Percentage increase rates in precision values of COVID-19 prediction for MLP-based models compared to precision values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for unique combinations

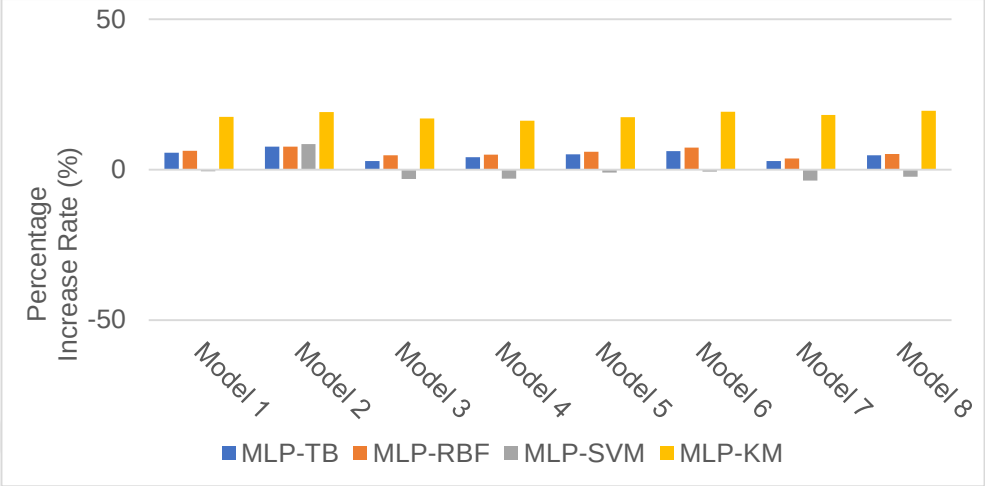


Figure 5.3. Percentage increase rates in recall values of COVID-19 prediction for MLP-based models compared to recall values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for unique combinations

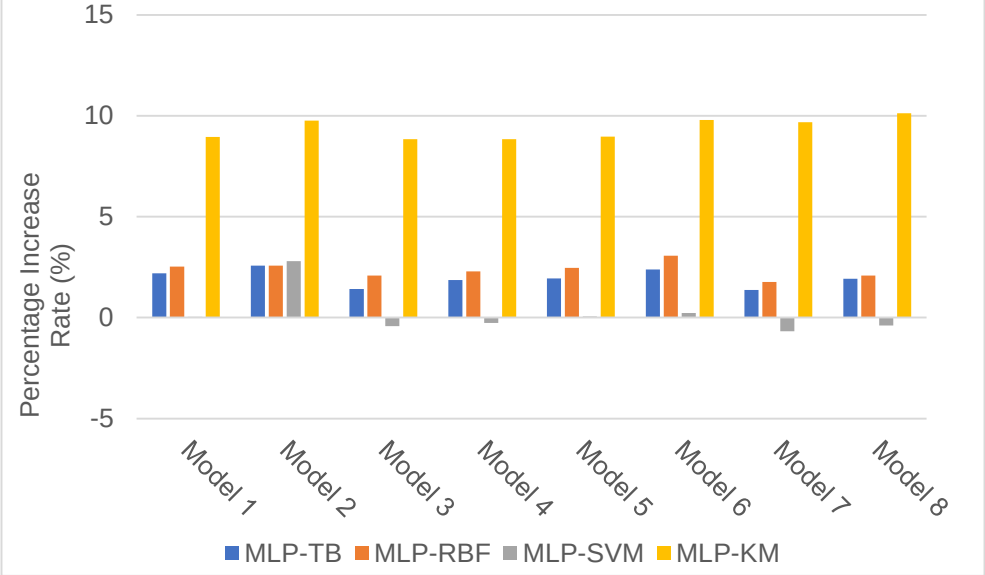


Figure 5.4. Percentage increase rates in F1-score values of COVID-19 prediction for MLP-based models compared to F1-score values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for unique combinations

5.2. Discussions for Models with Unique Combinations of Predictor Variables

The following comments regarding the results can be made for the prediction of COVID-19 developed using unique combinations of predictor variables:

- When the average accuracy results for each model using all machine learning classifiers are investigated, it is seen that Model 1 including the common variables with pregnant, and Model 5 including the common variables with renal comorbidity exhibit the same best performance with an accuracy value of 69.47%. In contrast, Model 2 including the common variables with race relatively produces the worst accuracy value of 69.15%.
- When the average precision results for each model using all machine learning classifiers are investigated, it is seen that Model 7 including the common variables with smoking comorbidity exhibits the best performance with a precision value of 65.81%. In contrast, Model 4 including the common variables with cardio comorbidity relatively produces the worst precision value of 65.54%.
- When the average recall results for each model using all machine learning classifiers are investigated, it is seen that Model 1 including the common variables with pregnant exhibits the best performance with a recall value of 81.50%. In contrast, Model 2 including the common variables with race relatively produces the worst recall value of 79.87%.
- When the average F1-score results for each model using all machine learning classifiers are investigated, it is seen that Model 1 including the common variables with pregnant exhibits the best performance with an F1-score value of 72.66%. In contrast, Model 2 including the common variables with race relatively produces the worst F1-score value of 72.06%.

- When the average accuracy and F1-score values of models including the unique combinations of predictor variables are compared, the general order of machine learning classifiers leading from the best to the worst results is MLP, SVM, TB, RBF, and KMC. When the average precision values of the models including the unique combinations of predictor variables are compared, the order of machine learning classifiers leading from the best to worst results is RBF, TB, MLP, SVM, and KMC. Finally, when the average recall values of the models including the unique combinations of predictor variables are compared, the order of machine learning classifiers leading from the best to worst results is SVM, MLP, TB, RBF, and KMC.
- It is observed that the MLP-based prediction models produce 0.75%, 0.93%, 0.39% and 4.45% higher accuracy values on average compared to those of TB-based, RBF-based, SVM-based, and KMC-based prediction models, respectively.
- It is observed that the MLP-based prediction models produce 0.52% and 0.88% higher precision values on average compared to that of SVM-based and KMC -based prediction models, respectively. In contrast, MLP-based prediction models produce 0.28% lower precision values compared to those of TB-based and RBF-based models.
- It is observed that the MLP-based prediction models produce 4.21%, 4.95%, and 15.50% higher recall values on average compared to those of TB-based, RBF-based, and KMC-based prediction models, respectively. In contrast, MLP-based prediction models produce 0.60% lower recall value on average compared to that of SVM-based prediction models.
- It is observed that the MLP-based prediction models produce 1.46%, 1.76%, 0.13%, and 6.98% higher F1-score values on average compared

to those of TB-based, RBF-based, SVM-based and KMC-based prediction models, respectively.

- Average training times of the MLP-based, TB-based, RBF-based, SVM-based, and KMC-based prediction models using computers with 3.20 GHz Dual Core CPU, 3 GB RAM, and 32-bit Windows 7 Professional OS are appr. 21 min, 2 min, 37 min, 11 hr., and 5 min, respectively.

5.3. Results for Prediction Models Using Double Combinations of the Predictor Variables

Table 5.5 through Table 5.8 show the accuracy, precision, recall, and F1-score values for COVID-19 prediction models including the double combinations of the predictor variables, respectively. Figure 5.5 through 5.8 illustrates the percentage increase rates in accuracy, precision, recall, and F1-score values for COVID-19 prediction models.

Table 5.5. Accuracy values for COVID-19 models including the double combinations of the predictor variables – Part 1

Models	Accuracy (%)				
	MLP	TB	RBF	SVM	KMC
Model 9	70.58	69.79	69.73	69.88	65.83
Model 10	70.70	69.95	69.71	70.12	66.33
Model 11	70.57	69.97	69.66	69.99	66.56
Model 12	70.68	69.92	69.67	70.38	66.67
Model 13	70.75	70.02	69.53	70.39	66.59
Model 14	70.71	70.13	69.84	70.32	66.33
Model 15	70.60	69.78	69.84	70.25	66.13
Model 16	70.63	69.92	69.77	69.59	65.86
Model 17	70.76	69.84	69.70	69.65	66.03
Model 18	70.62	69.89	69.65	69.73	66.06
Model 19	70.69	69.97	69.32	69.92	66.03
Model 20	70.58	70.00	69.56	69.89	66.05
Model 21	70.48	69.73	69.50	69.73	66.06
Model 22	70.75	70.03	69.80	69.89	66.20
Model 23	70.66	69.87	69.75	69.93	66.20

Table 5.6. Accuracy values for COVID-19 models including the double combinations of the predictor variables – Part 2

Models	Accuracy (%)				
	MLP	TB	RBF	SVM	KMC
Model 24	70.86	70.01	69.83	69.94	65.99
Model 25	70.62	70.08	69.91	70.07	66.59
Model 26	70.58	69.89	69.83	69.86	66.13
Model 27	70.70	69.94	69.69	69.88	66.30
Model 28	70.79	69.88	69.67	70.03	66.18
Model 29	70.73	69.97	69.57	69.84	65.97
Model 30	70.52	70.02	69.80	69.84	66.31
Model 31	70.77	69.85	69.79	70.36	66.17
Model 32	70.89	69.98	69.84	70.22	66.16
Model 33	70.70	69.95	69.81	70.23	66.01
Model 34	70.98	70.03	69.88	70.29	66.07
Model 35	70.67	69.89	69.69	70.30	65.97
Model 36	70.74	70.16	69.73	70.17	66.23

Table 5.7. Precision values for COVID-19 models including the double combinations of the predictor variables

Models	Precision (%)				
	MLP	TB	RBF	SVM	KMC
Model 9	65.95	66.20	66.21	66.55	64.72
Model 10	65.79	66.13	66.10	64.96	65.40
Model 11	65.39	66.12	66.07	64.79	65.13
Model 12	65.89	66.05	65.98	65.27	65.01
Model 13	66.18	66.27	65.99	65.15	65.36
Model 14	65.84	66.28	66.24	65.22	65.13
Model 15	65.71	65.82	66.15	65.03	64.94
Model 16	65.36	66.24	66.27	66.37	65.18
Model 17	65.37	66.18	66.29	66.59	65.00
Model 18	65.77	66.12	66.23	66.39	65.16
Model 19	65.86	66.16	65.96	66.70	65.05
Model 20	65.46	66.16	66.26	66.64	64.73
Model 21	65.98	66.24	66.11	66.50	64.68
Model 22	65.76	66.30	66.13	66.17	65.17
Model 23	66.00	66.07	66.04	66.05	65.01
Model 24	66.26	66.41	66.19	66.13	65.10
Model 25	65.24	66.30	66.29	64.87	65.99
Model 26	65.23	66.10	66.20	64.58	65.01
Model 27	65.96	66.14	65.99	64.69	65.37
Model 28	65.93	66.19	65.99	64.70	65.34
Model 29	65.79	66.20	66.04	64.54	64.61
Model 30	64.95	66.19	66.21	64.52	65.03
Model 31	65.65	66.11	66.09	65.13	65.17
Model 32	65.40	66.15	66.29	65.11	65.00
Model 33	65.79	66.04	66.15	64.99	65.16
Model 34	66.04	66.08	66.20	64.99	64.91
Model 35	65.87	66.13	66.08	64.97	65.14
Model 36	66.09	66.36	66.17	64.86	65.00

Table 5.8. Recall values for COVID-19 models including the double combinations of the predictor variables

Models	Recall (%)				
	MLP	TB	RBF	SVM	KMC
Model 9	85.07	80.87	80.59	79.91	69.59
Model 10	86.23	81.83	80.94	87.35	69.36
Model 11	87.42	81.90	80.80	87.59	71.28
Model 12	85.77	81.95	81.23	87.07	72.19
Model 13	84.88	81.53	80.62	87.68	70.58
Model 14	86.10	81.95	80.91	87.07	70.31
Model 15	86.15	82.29	81.27	87.60	70.13
Model 16	87.76	81.24	80.56	79.43	68.11
Model 17	88.29	81.13	80.17	78.87	69.45
Model 18	86.00	81.60	80.20	79.89	69.06
Model 19	85.94	81.74	79.83	79.57	69.31
Model 20	87.17	81.91	79.69	79.66	70.51
Model 21	84.56	80.48	80.04	79.49	70.80
Model 22	86.54	81.49	81.18	81.40	69.59
Model 23	85.21	81.69	81.34	82.00	70.17
Model 24	84.97	80.98	81.05	81.74	68.95
Model 25	88.26	81.68	81.01	87.56	68.43
Model 26	88.14	81.67	81.00	87.97	69.83
Model 27	85.58	81.71	81.25	87.55	69.32
Model 28	86.03	81.28	81.17	88.16	68.93
Model 29	86.37	81.62	80.56	88.02	70.66
Model 30	89.13	81.81	80.89	88.18	70.57
Model 31	87.13	81.47	81.27	87.66	69.50
Model 32	88.70	81.85	80.74	87.11	70.02
Model 33	86.26	82.14	81.17	87.68	68.83
Model 34	86.37	82.33	81.22	87.99	69.97
Model 35	85.83	81.54	80.93	88.13	68.73
Model 36	85.20	81.75	80.75	88.04	70.34

Table 5.9. F1-score values for COVID-19 models including the double combinations of the predictor variables

Models	F1-score (%)				
	MLP	TB	RBF	SVM	KMC
Model 9	74.30	72.80	72.70	72.62	67.07
Model 10	74.64	73.14	72.77	74.51	67.32
Model 11	74.81	73.17	72.70	74.48	68.07
Model 12	74.52	73.15	72.81	74.61	68.41
Model 13	74.37	73.11	72.58	74.75	67.87
Model 14	74.62	73.28	72.84	74.58	67.62
Model 15	74.56	73.14	72.94	74.65	67.43
Model 16	74.92	72.98	72.72	72.31	66.61
Model 17	75.12	72.90	72.58	72.21	67.15
Model 18	74.54	73.05	72.55	72.52	67.05
Model 19	74.57	73.13	72.24	72.57	67.11
Model 20	74.77	73.20	72.36	72.57	67.50
Model 21	74.12	72.67	72.41	72.42	67.60
Model 22	74.74	73.11	72.89	73.00	67.30
Model 23	74.38	73.05	72.90	73.17	67.49
Model 24	74.46	72.97	72.87	73.11	66.97
Model 25	75.03	73.19	72.91	74.53	67.19
Model 26	74.97	73.07	72.86	74.48	67.34
Model 27	74.50	73.11	72.83	74.41	67.29
Model 28	74.65	72.97	72.80	74.63	67.09
Model 29	74.69	73.10	72.58	74.48	67.50
Model 30	75.14	73.18	72.82	74.51	67.69
Model 31	74.88	72.99	72.90	74.73	67.26
Model 32	75.29	73.17	72.81	74.52	67.42
Model 33	74.65	73.22	72.89	74.65	66.94
Model 34	74.85	73.32	72.94	74.76	67.34
Model 35	74.53	73.03	72.76	74.80	66.89
Model 36	74.44	73.26	72.74	74.70	67.57

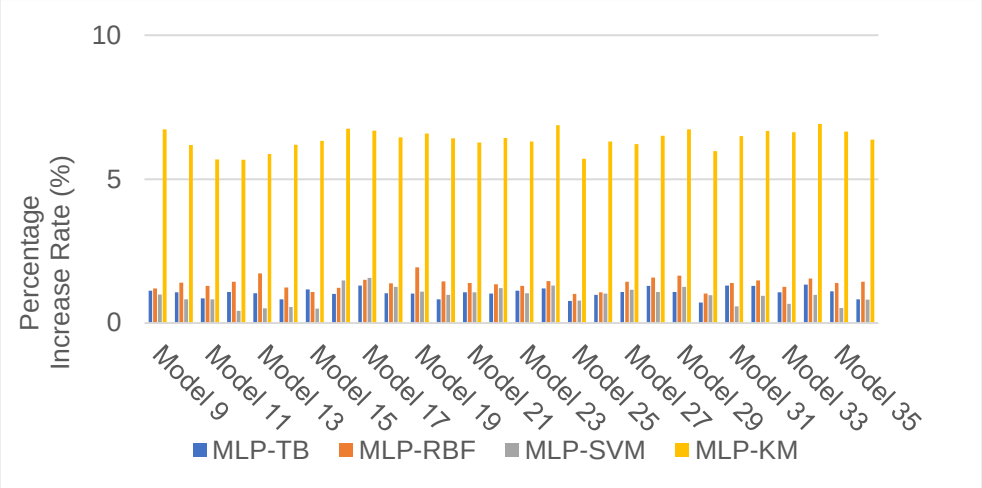


Figure 5.5. Percentage increase rates in accuracy values of COVID-19 prediction for MLP-based models compared to accuracy values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for double combinations

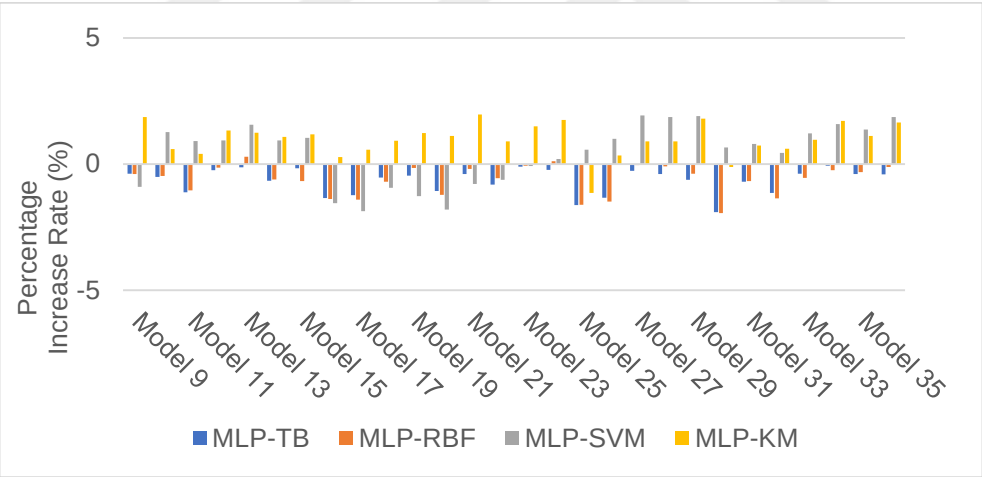


Figure 5.6. Percentage increase rates in precision values of COVID-19 prediction for MLP-based models compared to precision values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for double combinations

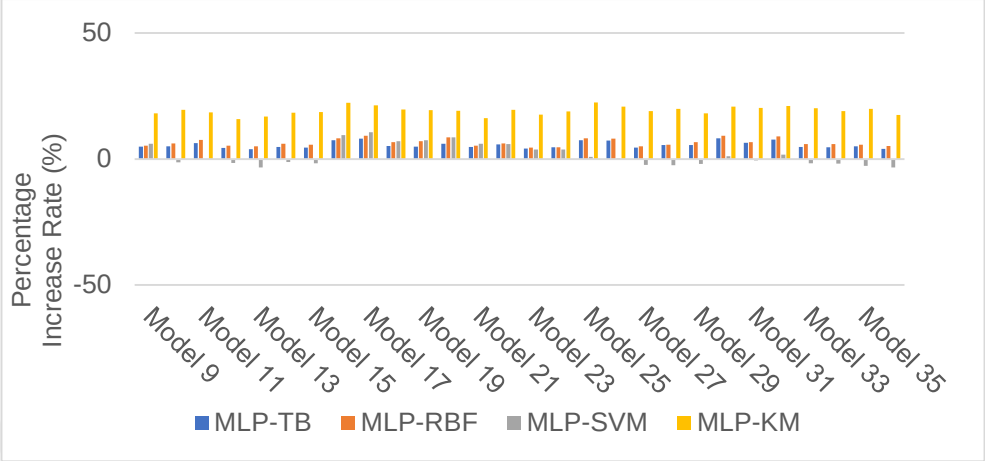


Figure 5.7. Percentage increase rates in recall values of COVID-19 prediction for MLP-based models compared to recall values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for double combinations

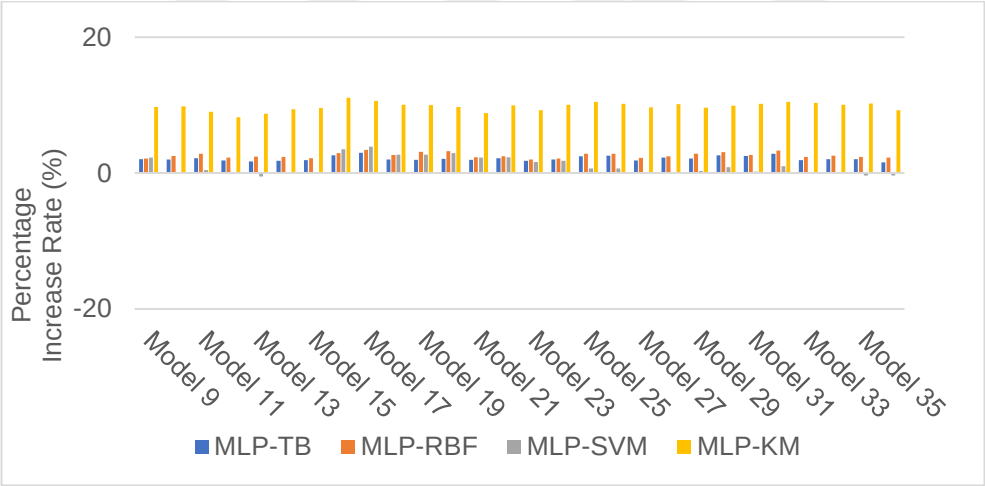


Figure 5.8. Percentage increase rates in F1-score values of COVID-19 prediction for MLP-based models compared to F1-score values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for double combinations

5.4. Discussion for Prediction Models Using Double Combinations of the Predictor Variables

The following comments regarding the results can be made for the prediction of COVID-19 developed using double combinations of predictor variables:

- When the average accuracy results for each model using all machine learning classifiers are investigated, it is seen that Model 14 including the common variables with pregnant and smoking comorbidity, exhibits the best performance with an accuracy value of 69.47%. In contrast, Model 21 including the common variables with race and obesity comorbidity relatively produces the worst accuracy value of 69.10%.
- When the average precision results for each model using all machine learning classifiers are investigated, it is seen that Model 24 including the common variables with lung and diabetes comorbidities exhibits the best performance with a precision value of 66.02%. In contrast, Model 30 including the common variables with cardio and obesity comorbidities relatively produces the worst precision value of 65.38%.
- When the average recall results for each model using all machine learning classifiers are investigated, it is seen that Model 30 including the common variables with cardio and obesity comorbidities exhibits the best performance with a recall value of 82.12%. In contrast, Model 21 including the common variables with race and obesity comorbidity relatively produces the worst recall value of 79.07%.
- When the average F1-score results for each model using all machine learning classifiers are investigated, it is seen that Model 12 including the common variables with pregnant and renal comorbidity exhibits the best performance with an F1-score value of 72.70%. In contrast, Model

21 including the common variables with race and obesity comorbidity relatively produces the worst F1-score value of 71.84%.

- When the average accuracy, recall and F1-score values of models including the double combinations of predictor variables are compared, the general order of machine learning classifiers leading from the best to the worst results is MLP, SVM, TB, RBF, and KMC. When the average precision values of the models including the double combinations of predictor variables are compared, the order of machine learning classifiers leading from the best to worst results is TB, RBF, MLP, SVM, and KMC.
- It is observed that the MLP-based prediction models produce 0.74%, 0.97%, 0.67% and 4.51% higher accuracy values on average compared to those of TB-based, RBF-based, SVM-based, and KMC-based prediction models, respectively.
- It is observed that the MLP-based prediction models produce 0.29% and 0.64% higher precision values on average compared to that of SVM-based and KMC -based prediction models, respectively. In contrast, MLP-based prediction models produce 0.44% and 0.41% lower precision values compared to those of TB-based and RBF-based models respectively.
- It is observed that the MLP-based prediction models produce 4.84%, 5.67%, 1.45%, and 16.66% higher recall values on average compared to those of TB-based, RBF-based, SVM-based, and KMC-based prediction models, respectively.
- It is observed that the MLP-based prediction models produce 1.59%, 1.94%, 0.78%, and 7.32% higher F1-score values on average compared to those of TB-based, RBF-based, SVM-based and KMC-based prediction models, respectively.

- Average training times of the MLP-based, TB-based, RBF-based, SVM-based, and KMC-based prediction models using computers with 3.20 GHz Dual Core CPU, 3 GB RAM, and 32-bit Windows 7 Professional OS are appr. 23 min, 2 min, 41 min, 10 hr., and 6 min, respectively.

5.5. Results for Models with Triple Combinations of Predictor Variables

Table 5.9 through Table 5.16 show the accuracy, precision, recall, and F1-score values for COVID-19 prediction models including the triple combinations of the predictor variables, respectively. Figure 5.9 through 5.12 illustrates the percentage increase rates in accuracy, precision, recall, and F1-score values for COVID-19 prediction models.

Table 5.10. Accuracy values for COVID-19 models including the triple combinations of the predictor variables – Part 1

Models	Accuracy (%)				
	MLP	TB	RBF	SVM	KMC
Model 37	70.63	69.80	69.75	69.62	65.92
Model 38	70.41	69.80	69.64	69.70	66.05
Model 39	70.58	69.79	69.84	69.75	66.22
Model 40	70.55	69.86	69.61	69.38	65.90
Model 41	70.67	69.90	69.86	69.91	66.16
Model 42	70.57	69.98	69.70	69.64	66.41
Model 43	70.73	69.95	69.69	69.83	66.44
Model 44	70.63	69.91	69.72	69.87	66.35
Model 45	70.58	70.05	69.66	69.94	66.32
Model 46	70.78	70.00	69.88	70.03	66.42
Model 47	70.72	69.95	69.64	69.91	66.35
Model 48	70.68	70.03	69.63	69.84	66.42
Model 49	70.73	70.04	69.69	69.97	66.24
Model 50	70.63	70.03	69.84	69.82	66.47
Model 51	70.57	69.85	69.67	69.75	66.38
Model 52	70.64	69.92	69.69	70.29	66.52
Model 53	70.83	69.99	69.76	70.20	66.30
Model 54	70.78	69.86	69.66	70.16	66.33
Model 55	70.83	69.92	69.65	70.28	66.20
Model 56	70.70	70.05	69.71	70.23	66.10
Model 57	70.65	69.94	69.99	70.12	66.51
Model 58	70.70	69.95	69.57	69.64	66.20
Model 59	70.62	69.83	69.73	69.60	65.98
Model 60	70.73	69.98	69.63	69.79	66.01
Model 61	70.62	70.01	69.74	69.69	65.77
Model 62	70.58	69.82	69.64	69.48	65.95
Model 63	70.55	69.82	69.62	69.63	66.03
Model 64	70.58	69.73	69.50	69.71	65.98

Table 5.11. Accuracy values for COVID-19 models including the triple combinations of the predictor variables – Part 2

Models	Accuracy (%)				
	MLP	TB	RBF	SVM	KMC
Model 65	70.62	69.98	69.73	69.62	66.33
Model 66	70.53	69.78	69.52	69.63	66.16
Model 67	70.54	69.98	69.77	69.88	66.09
Model 68	70.55	70.06	70.04	69.88	66.27
Model 69	70.52	69.76	69.58	69.62	66.16
Model 70	70.59	69.98	69.77	69.88	66.16
Model 71	70.66	70.02	69.55	69.59	65.80
Model 72	70.62	69.75	69.66	69.61	66.11
Model 73	70.65	70.06	69.64	69.83	66.36
Model 74	70.67	69.94	69.59	69.90	66.23
Model 75	70.71	70.11	69.84	69.83	66.25
Model 76	70.60	70.06	69.70	69.78	66.22
Model 77	70.84	69.96	69.67	69.95	66.28
Model 78	70.73	70.09	69.91	69.88	66.53
Model 79	70.67	69.92	69.72	69.95	66.25
Model 80	70.85	70.02	69.83	70.03	66.61
Model 81	70.66	70.03	69.77	69.94	66.25
Model 82	70.81	70.05	69.92	69.93	66.32
Model 83	70.70	70.02	69.49	69.88	66.24
Model 84	70.80	69.95	69.78	69.72	66.27
Model 85	70.66	69.86	69.68	69.78	66.44
Model 86	70.75	70.03	69.75	69.45	66.14
Model 87	70.58	69.99	69.49	69.77	66.31
Model 88	70.58	70.02	69.72	69.62	66.09
Model 89	70.69	70.06	69.69	70.20	66.09
Model 90	70.64	70.12	69.58	70.20	66.24
Model 91	70.66	69.99	69.80	70.03	66.18
Model 92	70.64	69.87	69.78	70.13	66.21

Table 5.12. Precision values for COVID-19 models including the triple combinations of the predictor variables – Part 1

Models	Precision (%)				
	MLP	TB	RBF	SVM	KMC
Model 37	65.87	66.22	66.25	66.41	64.49
Model 38	65.13	65.98	66.10	66.66	64.88
Model 39	65.74	66.17	66.22	66.44	65.17
Model 40	65.80	66.18	66.03	66.76	65.28
Model 41	65.05	66.14	66.34	66.66	64.77
Model 42	65.97	66.15	66.22	66.44	65.34
Model 43	65.60	66.16	66.12	66.15	65.36
Model 44	65.67	65.97	66.06	66.02	65.45
Model 45	65.10	66.27	66.09	64.67	65.50
Model 46	65.46	66.06	66.25	64.83	65.84
Model 47	65.96	66.09	66.02	66.17	65.79
Model 48	65.85	66.28	66.02	64.64	65.08
Model 49	65.64	66.19	66.10	64.63	65.12
Model 50	65.49	66.29	66.27	64.51	64.96
Model 51	64.84	65.90	66.07	64.42	65.32
Model 52	65.81	66.14	66.04	65.06	65.26
Model 53	65.13	66.20	66.16	65.10	65.52
Model 54	66.05	66.17	66.00	64.93	65.13
Model 55	66.16	66.04	66.10	64.96	65.00
Model 56	65.92	66.31	66.07	64.89	65.44
Model 57	65.26	66.20	66.37	64.80	65.17
Model 58	65.85	66.30	66.11	66.65	65.35
Model 59	66.06	66.20	66.29	66.37	65.02
Model 60	66.13	66.18	66.17	66.69	64.41
Model 61	65.73	66.43	66.27	66.52	64.96
Model 62	65.42	66.23	66.21	66.35	64.85
Model 63	64.52	66.13	66.17	66.57	64.87
Model 64	65.80	66.02	66.05	66.76	64.87

Table 5.13. Precision values for COVID-19 models including the triple combinations of the predictor variables – Part 2

Models	Precision (%)				
	MLP	TB	RBF	SVM	KMC
Model 65	65.51	66.42	66.26	66.68	65.07
Model 66	64.89	66.14	66.06	66.67	64.57
Model 67	65.63	66.25	66.23	66.70	65.17
Model 68	64.69	66.46	66.55	66.65	64.75
Model 69	65.39	66.15	66.09	66.42	64.99
Model 70	65.19	66.40	66.23	66.81	64.53
Model 71	64.78	66.44	66.11	66.51	64.99
Model 72	65.64	66.24	66.18	66.50	64.84
Model 73	65.67	66.21	66.04	66.15	65.15
Model 74	66.14	66.02	66.05	66.31	65.56
Model 75	66.09	66.26	66.30	66.14	65.29
Model 76	65.87	66.35	66.14	66.22	65.50
Model 77	66.07	66.12	66.14	64.66	65.36
Model 78	66.03	66.14	66.30	66.07	65.95
Model 79	65.52	66.22	66.14	66.23	65.30
Model 80	66.16	66.05	66.36	66.26	65.45
Model 81	64.91	66.28	66.22	66.23	65.39
Model 82	65.74	66.10	66.32	66.17	65.62
Model 83	66.12	66.21	65.96	64.56	65.36
Model 84	66.15	66.08	66.20	64.41	64.96
Model 85	65.83	66.13	66.11	64.43	65.24
Model 86	66.01	66.26	66.26	66.05	64.97
Model 87	64.71	66.32	65.96	64.34	65.58
Model 88	65.85	66.28	66.21	64.21	65.10
Model 89	65.81	66.28	66.14	64.89	65.17
Model 90	65.87	66.36	66.05	64.88	65.56
Model 91	65.31	66.12	66.26	64.73	65.29
Model 92	64.82	66.21	66.16	64.70	65.21

Table 5.14. Recall values for COVID-19 models including the triple combinations of the predictor variables – Part 1

Models	Recall (%)				
	MLP	TB	RBF	SVM	KMC
Model 37	85.63	80.85	80.50	79.39	70.88
Model 38	87.85	81.76	80.60	78.84	69.98
Model 39	85.98	80.99	80.97	79.78	69.70
Model 40	85.60	81.22	80.76	77.19	67.92
Model 41	89.32	81.55	80.64	79.65	70.84
Model 42	84.99	81.82	80.42	79.38	69.89
Model 43	87.15	81.71	80.78	81.21	69.96
Model 44	86.45	82.22	81.08	81.90	69.25
Model 45	88.73	81.63	80.73	87.88	68.98
Model 46	87.98	82.28	81.06	87.59	68.22
Model 47	85.64	81.91	80.93	81.46	68.13
Model 48	85.91	81.51	80.89	87.62	70.86
Model 49	86.99	81.92	80.83	88.23	69.95
Model 50	87.21	81.54	80.83	88.12	71.53
Model 51	89.88	82.26	80.90	88.21	69.83
Model 52	85.91	81.60	81.10	87.67	70.64
Model 53	89.66	81.68	80.89	87.11	68.79
Model 54	85.52	81.28	81.08	87.64	70.30
Model 55	85.26	82.02	80.69	88.06	70.23
Model 56	85.72	81.51	81.04	88.14	68.24
Model 57	88.31	81.46	81.06	88.05	70.94
Model 58	85.98	81.11	80.31	78.61	68.98
Model 59	84.78	81.03	80.26	79.46	69.18
Model 60	85.00	81.73	80.30	79.08	71.57
Model 61	86.15	80.91	80.39	79.25	68.48
Model 62	87.29	80.89	80.20	79.07	69.69
Model 63	91.31	81.26	80.27	78.83	69.90
Model 64	85.70	81.30	80.26	78.50	69.74

Table 5.15. Recall values for COVID-19 models including the triple combinations of the predictor variables – Part 2

Models	Recall (%)				
	MLP	TB	RBF	SVM	KMC
Model 65	87.11	80.81	80.41	78.42	70.50
Model 66	89.45	81.09	80.29	78.52	71.63
Model 67	86.26	81.49	80.66	79.41	69.11
Model 68	90.49	80.97	80.57	79.57	71.45
Model 69	87.19	80.95	80.41	79.36	70.07
Model 70	88.37	80.90	80.69	79.02	71.78
Model 71	90.53	80.91	80.20	78.90	68.48
Model 72	86.55	80.55	80.38	79.04	70.35
Model 73	86.54	81.92	80.85	81.21	70.36
Model 74	84.69	82.19	80.60	80.89	68.39
Model 75	85.05	81.93	80.73	81.24	69.42
Model 76	85.52	81.44	80.72	80.76	68.52
Model 77	85.68	81.86	80.60	87.95	69.26
Model 78	85.42	82.33	80.96	81.75	68.34
Model 79	87.29	81.31	80.81	81.43	69.39
Model 80	85.38	82.39	80.42	81.63	70.35
Model 81	89.94	81.53	80.71	81.38	69.03
Model 82	86.93	82.28	80.96	81.57	68.55
Model 83	84.91	81.76	80.56	88.15	69.11
Model 84	85.18	81.99	80.86	88.10	70.66
Model 85	85.91	81.44	80.75	88.30	70.39
Model 86	85.59	81.59	80.50	80.05	70.08
Model 87	90.54	81.24	80.56	88.69	68.66
Model 88	85.52	81.50	80.52	88.66	69.37
Model 89	86.16	81.68	80.69	88.04	69.09
Model 90	85.69	81.60	80.58	88.12	68.41
Model 91	88.15	82.03	80.71	88.02	69.10
Model 92	90.29	81.15	81.01	88.57	69.49

Table 5.16. F1-score values for COVID-19 models including the triple combinations of the predictor variables – Part 1

Models	F1-score (%)				
	MLP	TB	RBF	SVM	KMC
Model 37	74.46	72.81	72.68	72.32	67.53
Model 38	74.80	73.03	72.64	72.24	67.33
Model 39	74.51	72.83	72.86	72.50	67.36
Model 40	74.40	72.93	72.66	71.59	66.58
Model 41	75.28	73.04	72.80	72.58	67.67
Model 42	74.28	73.16	72.63	72.34	67.54
Model 43	74.86	73.12	72.72	72.91	67.58
Model 44	74.64	73.20	72.81	73.11	67.30
Model 45	75.10	73.15	72.68	74.51	67.19
Model 46	75.07	73.28	72.91	74.51	67.01
Model 47	74.52	73.16	72.72	73.03	66.94
Model 48	74.56	73.11	72.70	74.40	67.85
Model 49	74.82	73.22	72.73	74.61	67.45
Model 50	74.81	73.13	72.83	74.49	68.08
Model 51	75.33	73.18	72.74	74.46	67.50
Model 52	74.53	73.06	72.80	74.69	67.84
Model 53	75.45	73.13	72.79	74.51	67.12
Model 54	74.54	72.95	72.77	74.60	67.62
Model 55	74.51	73.17	72.67	74.77	67.51
Model 56	74.53	73.13	72.79	74.75	66.81
Model 57	75.06	73.04	72.98	74.66	67.93
Model 58	74.58	72.96	72.52	72.14	67.12
Model 59	74.26	72.87	72.61	72.33	67.03
Model 60	74.39	73.14	72.55	72.36	67.80
Model 61	74.57	72.96	72.65	72.33	66.67
Model 62	74.79	72.83	72.54	72.15	67.18
Model 63	75.61	72.92	72.54	72.19	67.29
Model 64	74.44	72.87	72.47	72.16	67.22

Table 5.17. F1-score values for COVID-19 models including the triple combinations of the predictor variables – Part 2

Models	F1-score (%)				
	MLP	TB	RBF	SVM	K-Means
Model 65	74.78	72.91	72.65	72.07	67.68
Model 66	75.22	72.85	72.48	72.11	67.92
Model 67	74.54	73.08	72.74	72.50	67.08
Model 68	75.45	73.00	72.89	72.54	67.93
Model 69	74.73	72.80	72.55	72.32	67.44
Model 70	75.03	72.94	72.75	72.40	67.96
Model 71	75.52	72.96	72.48	72.18	66.69
Model 72	74.66	72.70	72.59	72.23	67.49
Model 73	74.67	73.23	72.70	72.91	67.65
Model 74	74.28	73.22	72.61	72.88	66.94
Model 75	74.38	73.27	72.81	72.92	67.29
Model 76	74.42	73.12	72.71	72.77	66.98
Model 77	74.61	73.15	72.66	74.53	67.26
Model 78	74.48	73.36	72.90	73.08	67.12
Model 79	74.85	72.99	72.74	73.05	67.28
Model 80	74.55	73.32	72.72	73.15	67.81
Model 81	75.40	73.12	72.75	73.03	67.16
Model 82	74.86	73.31	72.91	73.07	67.05
Model 83	74.35	73.17	72.53	74.53	67.18
Model 84	74.47	73.18	72.80	74.42	67.69
Model 85	74.54	72.99	72.70	74.50	67.72
Model 86	74.53	73.13	72.69	72.38	67.43
Model 87	75.48	73.02	72.53	74.58	67.09
Model 88	74.40	73.10	72.67	74.48	67.17
Model 89	74.62	73.18	72.70	74.71	67.07
Model 90	74.48	73.20	72.59	74.73	66.96
Model 91	75.03	73.22	72.77	74.60	67.14
Model 92	75.46	72.92	72.83	74.78	67.28

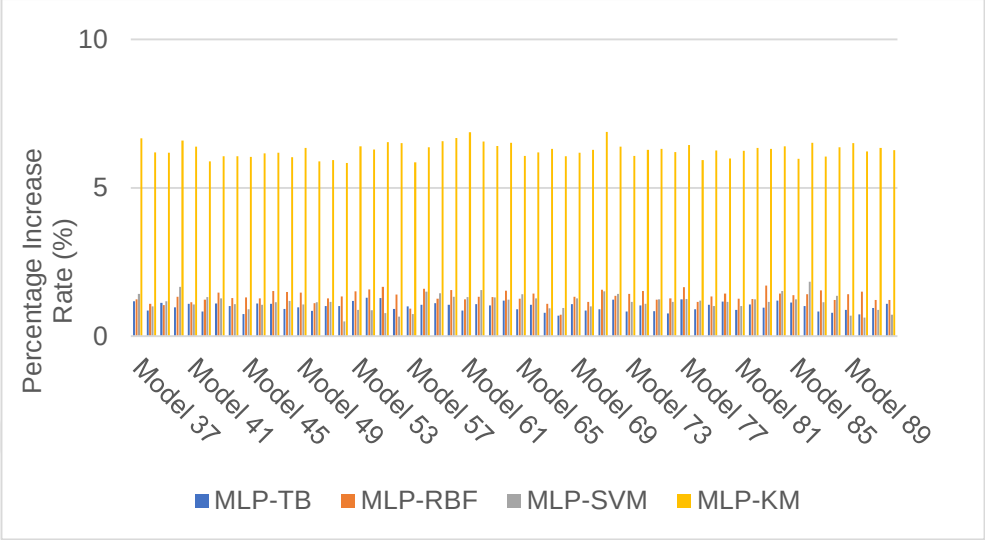


Figure 5.9. Percentage increase rates in accuracy values of COVID-19 prediction for MLP-based models compared to accuracy values obtained by TB-, RBF-SVM-, and KMC-based models for triple combinations

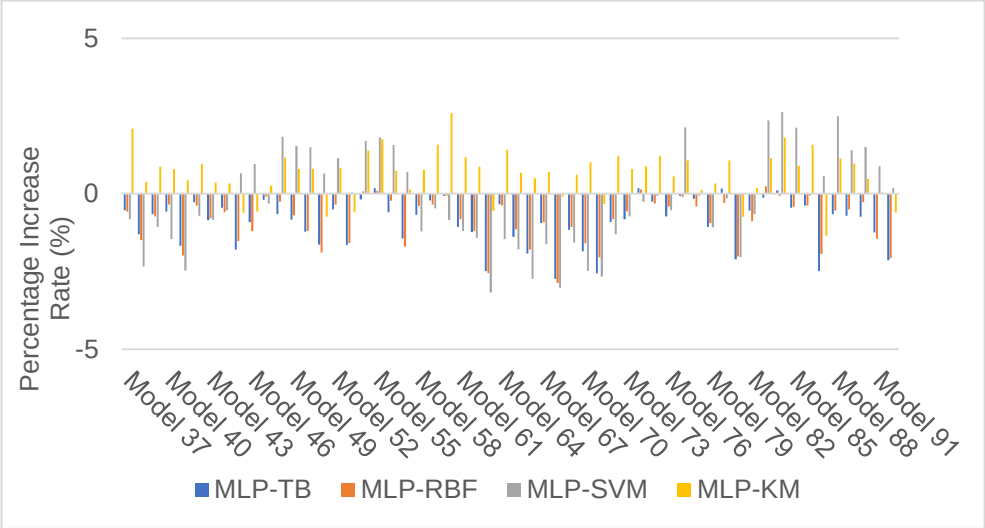


Figure 5.10. Percentage increase rates in precision values of COVID-19 prediction for MLP-based models compared to precision values obtained by TB-RBF-, SVM-, and KMC-based models for triple combinations

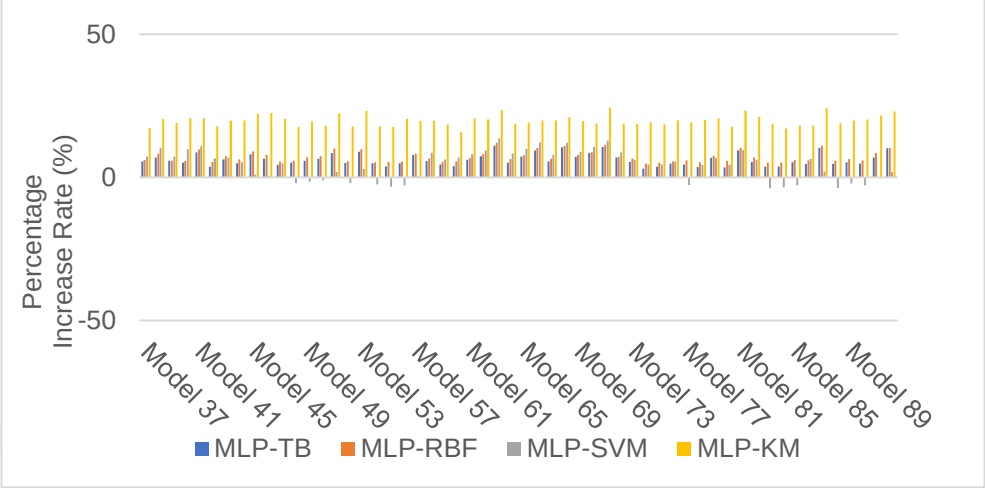


Figure 5.11. Percentage increase rates in recall values of COVID-19 prediction for MLP-based models compared to recall values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for triple combinations

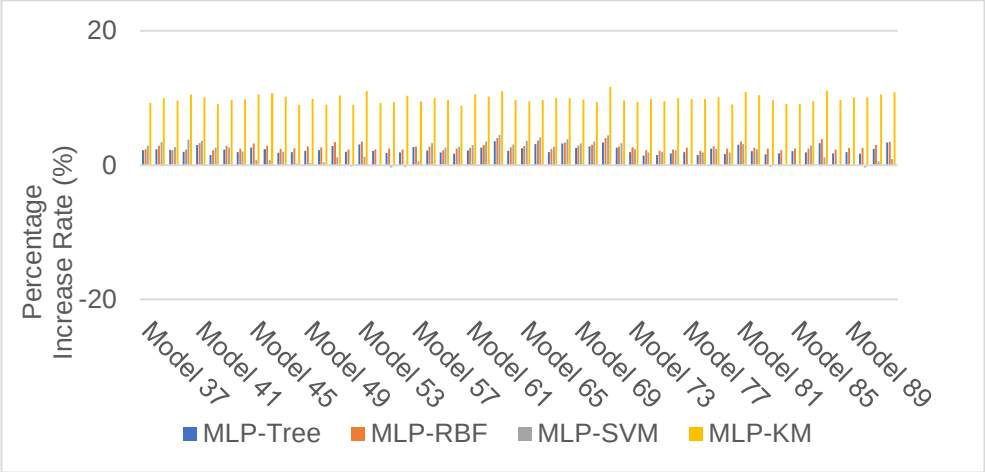


Figure 5.12. Percentage increase rates in F1-score values of COVID-19 prediction for MLP-based models compared to F1-score values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for triple combinations

5.6. Discussions for Models with Triple Combinations of Predictor Variables

The following comments regarding the results can be made for the prediction of COVID-19 developed using triple combinations of predictor variables:

- When the average accuracy results for each model using all machine learning classifiers are investigated, it is seen that Model 80 including the common variables with lung, diabetes and smoking comorbidities, exhibits the best performance with an accuracy value of 69.47%. In contrast, Model 40 including the common variables with pregnant, race, and diabetes comorbidity relatively produces the worst accuracy value of 69.06%.
- When the average precision results for each model using all machine learning classifiers are investigated, it is seen that Model 78 including the common variables with lung, renal, and smoking comorbidities exhibits the best performance with a precision value of 66.10%. In contrast, Model 51 including the common variables with pregnant, cardio, and obesity comorbidities relatively produces the worst precision value of 65.31%.
- When the average recall results for each model using all machine learning classifiers are investigated, it is seen that Model 51 including the common variables with pregnant, cardio, and obesity comorbidities exhibits the best performance with a recall value of 82.22%. In contrast, Model 40 including the common variables with pregnant, race, and diabetes comorbidity relatively produces the worst recall value of 78.54%.
- When the average F1-score results for each model using all machine learning classifiers are investigated, it is seen that Model 57 including the common variables with pregnant, smoking, and obesity comorbidities exhibits the best performance with a F1-score value of

72.73%. In contrast, Model 40 including the common variables with pregnant, race, and diabetes comorbidity relatively produces the worst F1-score value of 71.63%.

- When the average accuracy values of models including the triple combinations of predictor variables are compared, the general order of machine learning classifiers leading from the best to the worst results is MLP, TB, SVM, RBF, and KMC. When the average precision values of the models including the triple combinations of predictor variables are compared, the order of machine learning classifiers leading from the best to worst results is TB, RBF, SVM, MLP, and KMC. Finally, when the average recall and F1-score values of the models including the triple combinations of predictor variables are compared, the order of machine learning classifiers leading from the best to worst results is MLP, SVM, TB, RBF, and KMC.
- It is observed that the MLP-based prediction models produce 0.81%, 0.95%, 4.43% and 0.71% higher accuracy values on average compared to those of SVM-based, RBF-based, K-Means-based, and TB-based prediction models, respectively.
- It is observed that the MLP-based prediction models produce 0.42% higher precision value on average compared to those of K-Means-based prediction models, respectively. In contrast, it is observed that the MLP-based prediction models produce 0.15%, 0.56%, and 0.60% lower precision values compared to those of SVM-based, RBF-based, and TB-based models, respectively.
- It is observed that the MLP-based prediction models produce 3.85%, 6.26%, 17.26% and 5.40% higher recall values on average compared to those of SVM-based, RBF-based, K-Means-based, and TB-based prediction models, respectively.

- It is observed that the MLP-based prediction models produce 1.43%, 2.06%, 7.41%, and 1.69% higher F1-score values on average compared to those of SVM-based, RBF-based, K-Means-based and TB-based prediction models, respectively.
- Average training times of the MLP-based, TB-based, RBF-based, SVM-based, and KMC-based prediction models using computers with 3.20 GHz Dual Core CPU, 3 GB RAM, and 32-bit Windows 7 Professional OS are appr. 22 min, 2 min, 47 min, 10 hr, and 6 min, respectively.

5.7. Results for Prediction Models Using Quadruple Combinations of the Predictor Variables

Table 5.17 through Table 5.28 show the accuracy, precision, recall, and F1-score values for COVID-19 prediction models including the quadruple combinations of the predictor variables, respectively. Figure 5.13 through 5.16 illustrates the percentage increase rates in accuracy, precision, recall, and F1-score values for COVID-19 prediction models.

Table 5.18. Accuracy values for COVID-19 models including the quadruple combinations of the predictor variables – Part 1

Models	Accuracy (%)				
	MLP	TB	RBF	SVM	KMC
Model 93	70.56	69.79	69.83	69.63	66.16
Model 94	70.67	69.88	69.77	69.61	66.12
Model 95	70.61	70.04	69.73	69.75	66.29
Model 96	70.77	69.94	69.83	69.71	65.56
Model 97	70.66	69.88	69.58	69.55	65.75
Model 98	70.58	69.83	69.73	69.60	65.95
Model 99	70.61	69.83	69.60	69.71	66.01
Model 100	70.75	69.89	69.77	69.64	66.36
Model 101	70.42	69.78	69.50	69.64	66.11
Model 102	70.67	69.95	69.74	69.86	66.25
Model 103	70.75	69.92	69.73	69.88	66.45
Model 104	70.58	69.77	69.64	69.63	66.42
Model 105	70.58	69.84	69.58	69.89	66.13
Model 106	70.62	69.88	69.50	69.64	65.75
Model 107	70.56	69.87	69.56	69.67	66.12
Model 108	70.81	70.04	69.77	69.77	66.64
Model 109	70.47	70.06	69.73	69.58	66.38
Model 110	70.64	70.10	69.90	69.77	66.47
Model 111	70.58	69.94	69.83	69.79	66.50
Model 112	70.79	70.04	69.72	69.87	66.41
Model 113	70.67	70.07	69.80	69.89	66.33
Model 114	70.71	69.86	69.66	69.94	66.25
Model 115	70.61	70.10	69.89	70.01	66.43
Model 116	70.75	70.08	69.80	69.96	66.16
Model 117	70.77	70.05	69.88	69.98	66.36
Model 118	70.64	69.98	69.58	69.83	66.23
Model 119	70.70	69.89	69.81	69.69	66.14

Table 5.19. Accuracy values for COVID-19 models including the quadruple combinations of the predictor variables – Part 2

Models	Accuracy (%)				
	MLP	TB	RBF	SVM	KMC
Model 28	70.70	69.92	69.49	69.66	66.41
Model 29	70.65	69.99	69.72	69.65	66.16
Model 30	70.56	69.92	69.62	69.66	66.11
Model 31	70.67	69.94	69.64	69.56	66.31
Model 32	70.84	69.89	69.73	70.16	66.51
Model 33	70.73	70.03	69.86	70.13	66.21
Model 34	70.78	70.09	69.84	70.00	66.33
Model 35	70.75	70.03	69.78	69.38	66.22
Model 36	70.56	69.99	69.77	69.60	66.28
Model 37	70.55	69.94	69.91	69.61	65.86
Model 38	70.67	69.88	69.69	69.42	66.40
Model 39	70.52	69.83	69.67	69.56	66.14
Model 40	70.57	69.95	70.03	69.81	66.08
Model 41	70.64	69.88	69.97	69.63	66.09
Model 42	70.60	69.95	69.73	69.53	66.15
Model 43	70.64	70.15	69.84	69.70	66.47
Model 44	70.55	69.97	69.73	69.59	66.09
Model 45	70.52	70.02	69.99	69.41	66.24
Model 46	70.58	69.84	69.87	69.66	65.81
Model 47	70.66	69.89	69.75	69.56	66.20
Model 48	70.58	69.71	69.60	69.60	66.16
Model 49	70.66	69.88	69.74	69.67	66.18
Model 50	70.45	69.84	69.49	69.64	66.24
Model 51	70.74	69.84	69.69	69.45	66.59
Model 52	70.53	69.89	69.83	69.80	66.11
Model 53	70.67	69.85	69.63	69.66	65.75
Model 54	70.64	69.88	69.71	69.57	66.11
Model 55	70.68	69.92	69.77	69.61	66.17

Table 5.20. Accuracy values for COVID-19 models including the quadruple combinations of the predictor variables – Part 3

Models	Accuracy (%)				
	MLP	**TB**	**RBF**	**SVM**	**KMC**
Model 148	70.67	70.23	69.68	69.81	66.23
Model 149	70.72	70.02	69.77	69.80	66.47
Model 150	70.62	70.00	69.59	69.73	66.25
Model 151	70.66	70.27	69.71	69.92	66.31
Model 152	70.70	70.14	69.69	69.73	66.14
Model 153	70.66	70.08	69.81	69.67	66.02
Model 154	70.69	70.08	69.91	69.94	66.52
Model 155	70.69	69.94	70.03	69.91	66.60
Model 156	70.81	70.03	69.97	69.95	66.40
Model 157	70.69	69.92	69.83	70.07	66.27
Model 158	70.79	70.07	69.75	69.58	66.20
Model 159	70.66	69.98	69.65	69.69	66.39
Model 160	70.69	69.92	69.80	69.52	66.22
Model 161	70.78	69.91	69.73	69.52	65.86
Model 162	70.80	70.02	69.96	69.98	66.01

Table 5.21. Precision values for COVID-19 models including the quadruple combinations of the predictor variables – Part 1

Models	Precision (%)				
	MLP	**TB**	**RBF**	**SVM**	**KMC**
Model 93	64.75	66.19	66.25	66.66	64.49
Model 94	64.86	66.12	66.20	66.39	65.24
Model 95	65.69	66.17	66.21	66.69	64.77
Model 96	65.66	66.26	66.26	66.55	64.17
Model 97	65.89	66.27	66.10	66.43	65.03
Model 98	65.03	65.94	66.18	66.57	64.70
Model 99	64.79	65.93	66.19	66.78	64.68
Model 100	64.96	66.26	66.23	66.71	64.83
Model 101	64.89	66.06	66.04	66.70	64.68
Model 102	65.48	66.18	66.18	66.70	65.33
Model 103	65.95	66.32	66.23	66.67	64.99
Model 104	64.89	66.21	66.05	66.45	65.10
Model 105	64.81	66.20	66.14	66.83	64.50
Model 106	65.14	66.35	66.01	66.60	64.74
Model 107	65.55	66.12	66.13	66.55	64.73
Model 108	65.74	66.22	66.19	66.14	65.54
Model 109	64.77	66.12	66.17	64.18	65.24
Model 110	65.42	66.39	66.32	66.12	65.67
Model 111	65.53	66.01	66.26	66.25	65.29
Model 112	65.81	66.27	66.09	66.14	65.46
Model 113	65.60	66.09	66.23	66.08	65.64
Model 114	65.88	66.10	66.10	66.23	65.32
Model 115	65.79	66.25	66.30	66.26	65.47
Model 116	66.08	66.35	66.24	66.27	65.23
Model 117	65.94	66.20	66.28	66.22	65.71
Model 118	65.39	66.23	66.06	64.48	65.03
Model 119	65.78	66.17	66.29	64.37	65.17

Table 5.22. Precision values for COVID-19 models including the quadruple combinations of the predictor variables – Part 2

Models	Precision (%)				
	MLP	TB	RBF	SVM	KMC
Model 120	65.40	66.20	65.98	64.33	65.24
Model 121	65.36	66.25	66.17	64.23	64.97
Model 122	64.68	66.25	66.06	64.23	65.33
Model 123	66.14	66.20	66.15	64.13	65.20
Model 124	65.57	66.21	66.14	64.82	65.31
Model 125	65.86	66.27	66.23	64.80	65.59
Model 126	65.55	66.12	66.23	64.69	65.06
Model 127	65.68	66.38	66.21	66.76	65.28
Model 128	65.34	66.31	66.26	66.64	65.41
Model 129	65.59	66.40	66.41	66.78	65.02
Model 130	66.16	66.37	66.24	66.55	65.37
Model 131	65.46	66.25	66.18	66.66	65.02
Model 132	65.08	66.17	66.42	66.74	64.43
Model 133	64.80	66.25	66.41	66.50	64.70
Model 134	64.67	66.38	66.22	66.40	65.08
Model 135	65.96	66.60	66.33	66.69	65.07
Model 136	65.77	66.24	66.29	66.58	64.85
Model 137	64.69	66.26	66.45	66.36	65.08
Model 138	65.72	66.17	66.26	66.77	64.65
Model 139	65.55	66.12	66.23	66.64	64.99
Model 140	65.73	66.17	66.11	66.66	64.51
Model 141	65.99	66.37	66.26	66.84	65.14
Model 142	64.85	66.34	65.99	66.78	65.27
Model 143	66.05	66.29	66.30	66.62	65.00
Model 144	65.72	66.38	66.37	66.75	64.45
Model 145	65.87	66.10	66.13	66.59	65.00
Model 146	65.11	66.33	66.27	66.48	65.06
Model 147	65.90	66.22	66.32	66.67	65.22

Table 5.23. Precision values for COVID-19 models including the quadruple combinations of the predictor variables – Part 3

Models	Precision (%)				
	MLP	TB	RBF	SVM	KMC
Model 148	65.99	66.37	66.07	66.29	65.35
Model 149	65.83	66.13	66.24	66.17	65.26
Model 150	65.75	66.00	66.12	66.21	65.31
Model 151	64.94	66.53	66.19	66.37	65.34
Model 152	66.06	66.39	66.22	66.27	65.18
Model 153	65.90	66.27	66.25	66.15	64.97
Model 154	65.17	65.98	66.36	66.23	65.38
Model 155	65.56	66.09	66.40	66.23	65.59
Model 156	65.81	66.17	66.34	66.23	65.52
Model 157	65.11	66.30	66.26	66.37	65.55
Model 158	65.51	66.34	66.21	64.18	64.99
Model 159	65.81	66.22	66.08	64.26	65.64
Model 160	65.50	66.09	66.25	64.11	65.27
Model 161	65.99	66.25	66.24	64.02	64.80
Model 162	66.04	66.12	66.35	64.58	64.99

Table 5.24. Recall values for COVID-19 models including the quadruple combinations of the predictor variables – Part 1

Models	Recall (%)				
	MLP	TB	RBF	SVM	KMC
Model 93	90.25	80.92	80.82	78.51	71.91
Model 94	90.21	81.57	80.81	79.42	69.02
Model 95	86.27	82.00	80.57	78.93	71.43
Model 96	87.08	81.25	80.78	79.26	70.47
Model 97	85.63	80.99	80.37	79.06	68.16
Model 98	89.04	82.03	80.71	78.73	70.22
Model 99	90.27	82.07	80.15	78.44	70.53
Model 100	90.10	81.07	80.69	78.39	71.51
Model 101	88.95	81.38	80.32	78.46	70.99
Model 102	87.46	81.61	80.73	79.29	69.25
Model 103	85.79	80.96	80.52	79.52	71.34
Model 104	89.70	80.76	80.82	79.31	70.80
Model 105	90.09	81.10	80.21	78.98	71.73
Model 106	88.72	80.69	80.39	78.83	69.17
Model 107	86.67	81.49	80.20	79.10	70.83
Model 108	86.93	81.81	80.84	81.04	70.17
Model 109	89.81	82.32	80.72	88.62	70.11
Model 110	87.56	81.41	80.85	81.12	69.03
Model 111	86.81	82.18	80.79	80.67	70.42
Model 112	86.53	81.63	80.98	81.41	69.47
Model 113	86.93	82.44	80.78	81.72	68.55
Model 114	85.93	81.52	80.73	81.34	69.28
Model 115	85.86	81.98	80.94	81.55	69.52
Model 116	85.30	81.47	80.75	81.30	69.19
Model 117	85.89	81.92	80.94	81.56	68.42
Model 118	87.70	81.54	80.53	88.26	70.22
Model 119	86.32	81.39	80.64	88.22	69.34

Table 5.25. Recall values for COVID-19 models including the quadruple combinations of the predictor variables – Part 2

Models	Recall (%)				
	MLP	TB	RBF	SVM	KMC
Model 120	87.89	81.42	80.51	88.28	70.24
Model 121	87.85	81.51	80.68	88.69	70.15
Model 122	90.60	81.23	80.71	88.73	68.69
Model 123	84.72	81.48	80.41	88.74	69.97
Model 124	87.75	81.26	80.87	88.14	70.40
Model 125	86.10	81.59	81.04	88.10	68.19
Model 126	87.63	82.42	80.97	88.10	70.56
Model 127	86.93	81.17	80.83	77.19	69.33
Model 128	87.61	81.31	80.55	78.49	69.10
Model 129	86.46	80.72	80.57	78.05	68.67
Model 130	84.61	80.57	80.35	78.07	69.75
Model 131	86.86	80.84	80.46	78.25	69.87
Model 132	88.78	81.63	81.01	78.99	71.78
Model 133	90.37	81.08	80.81	79.11	70.83
Model 134	90.83	80.84	80.58	79.06	69.69
Model 135	85.30	80.86	80.58	78.72	71.10
Model 136	85.69	81.43	80.25	78.67	70.26
Model 137	90.36	81.55	80.76	78.74	70.10
Model 138	86.02	81.21	80.97	78.27	69.77
Model 139	87.11	81.59	80.58	78.31	70.25
Model 140	85.96	80.67	80.44	78.45	71.86
Model 141	85.23	80.61	80.45	78.09	69.62
Model 142	89.32	80.55	80.43	78.14	69.42
Model 143	85.34	80.74	80.10	77.99	71.90
Model 144	85.83	80.60	80.37	78.88	71.86
Model 145	85.81	81.51	80.48	78.93	68.26
Model 146	88.92	80.76	80.29	78.96	69.59
Model 147	85.71	81.35	80.33	78.43	69.30

Table 5.26. Recall values for COVID-19 models including the quadruple combinations of the predictor variables – Part 3

Models	Recall (%)				
	MLP	TB	RBF	SVM	KMC
Model 148	85.28	82.02	80.90	80.65	69.08
Model 149	86.18	82.08	80.64	81.02	70.46
Model 150	86.08	82.50	80.37	80.56	69.31
Model 151	89.80	81.56	80.58	80.75	69.47
Model 152	85.16	81.58	80.42	80.36	69.34
Model 153	85.64	81.80	80.76	80.57	69.49
Model 154	88.87	82.88	80.74	81.38	70.20
Model 155	87.17	81.88	81.10	81.24	69.84
Model 156	86.64	81.93	81.04	81.44	69.23
Model 157	89.13	81.03	80.79	81.38	68.59
Model 158	87.82	81.48	80.69	88.65	70.25
Model 159	85.98	81.59	80.75	88.74	68.79
Model 160	87.44	81.80	80.74	88.65	69.32
Model 161	85.77	81.15	80.46	89.13	69.46
Model 162	85.64	82.11	81.01	88.49	69.42

Table 5.27. F1-score values for COVID-19 models including the quadruple combinations of the predictor variables – Part 1

Models	F1-score (%)				
	MLP	TB	RBF	SVM	KMC
Model 93	75.40	72.82	72.81	72.10	68.00
Model 94	75.46	73.04	72.78	72.32	67.07
Model 95	74.59	73.24	72.69	72.29	67.93
Model 96	74.87	72.99	72.80	72.35	67.18
Model 97	74.48	72.90	72.54	72.19	66.56
Model 98	75.16	73.11	72.73	72.14	67.35
Model 99	75.44	73.12	72.50	72.14	67.48
Model 100	75.49	72.92	72.75	72.08	68.00
Model 101	75.04	72.92	72.48	72.10	67.69
Model 102	74.89	73.09	72.74	72.45	67.23
Model 103	74.57	72.91	72.68	72.53	68.02
Model 104	75.30	72.77	72.69	72.31	67.83
Model 105	75.39	72.90	72.50	72.40	67.92
Model 106	75.12	72.82	72.50	72.20	66.88
Model 107	74.64	73.01	72.49	72.28	67.64
Model 108	74.87	73.19	72.78	72.84	67.77
Model 109	75.26	73.33	72.73	74.45	67.59
Model 110	74.89	73.14	72.87	72.85	67.31
Model 111	74.68	73.21	72.81	72.75	67.76
Model 112	74.76	73.15	72.78	72.99	67.41
Model 113	74.78	73.36	72.79	73.08	67.06
Model 114	74.58	73.00	72.69	73.01	67.24
Model 115	74.50	73.28	72.89	73.11	67.44
Model 116	74.47	73.14	72.78	73.02	67.15
Model 117	74.61	73.22	72.88	73.10	67.04
Model 118	74.92	73.09	72.58	74.52	67.53
Model 119	74.66	73.00	72.76	74.43	67.19

Table 5.28. F1-score values for COVID-19 models including the quadruple combinations of the predictor variables – Part 2

Models	F1-score (%)				
	MLP	TB	RBF	SVM	KMC
Model 120	75.00	73.03	72.52	74.43	67.65
Model 121	74.96	73.09	72.71	74.50	67.46
Model 122	75.47	72.98	72.65	74.52	66.97
Model 123	74.28	73.05	72.59	74.46	67.50
Model 124	75.05	72.97	72.76	74.70	67.76
Model 125	74.63	73.14	72.89	74.68	66.87
Model 126	75.00	73.37	72.86	74.60	67.70
Model 127	74.83	73.03	72.79	71.59	67.24
Model 128	74.85	73.04	72.71	72.08	67.20
Model 129	74.59	72.86	72.81	71.98	66.80
Model 130	74.26	72.79	72.61	71.85	67.49
Model 131	74.66	72.82	72.63	71.99	67.36
Model 132	75.10	73.09	73.00	72.35	67.91
Model 133	75.48	72.92	72.91	72.26	67.62
Model 134	75.55	72.90	72.70	72.18	67.31
Model 135	74.39	73.04	72.77	72.21	67.95
Model 136	74.42	73.05	72.61	72.12	67.44
Model 137	75.40	73.12	72.91	72.02	67.49
Model 138	74.51	72.92	72.88	72.07	67.11
Model 139	74.80	73.04	72.71	72.01	67.52
Model 140	74.50	72.70	72.58	72.07	67.98
Model 141	74.39	72.80	72.67	72.03	67.30
Model 142	75.14	72.76	72.50	72.02	67.28
Model 143	74.47	72.80	72.55	71.86	68.28
Model 144	74.44	72.80	72.70	72.31	67.95
Model 145	74.53	73.00	72.60	72.23	66.59
Model 146	75.18	72.84	72.61	72.18	67.25
Model 147	74.51	73.01	72.66	72.07	67.20

Table 5.29. F1-score values for COVID-19 models including the quadruple combinations of the predictor variables – Part 3

Models	F1-score (%)				
	MLP	TB	RBF	SVM	KMC
Model 148	74.41	73.37	72.74	72.77	67.17
Model 149	74.64	73.25	72.74	72.85	67.76
Model 150	74.55	73.33	72.55	72.68	67.25
Model 151	75.37	73.28	72.68	72.86	67.34
Model 152	74.40	73.21	72.63	72.64	67.19
Model 153	74.48	73.22	72.79	72.65	67.16
Model 154	75.20	73.47	72.85	73.03	67.71
Model 155	74.83	73.14	73.02	72.97	67.65
Model 156	74.80	73.21	72.96	73.05	67.32
Model 157	75.25	72.93	72.81	73.11	67.03
Model 158	75.04	73.14	72.74	74.46	67.52
Model 159	74.56	73.11	72.68	74.54	67.18
Model 160	74.90	73.11	72.78	74.41	67.23
Model 161	74.59	72.95	72.66	74.51	67.05
Model 162	74.57	73.25	72.95	74.67	67.13

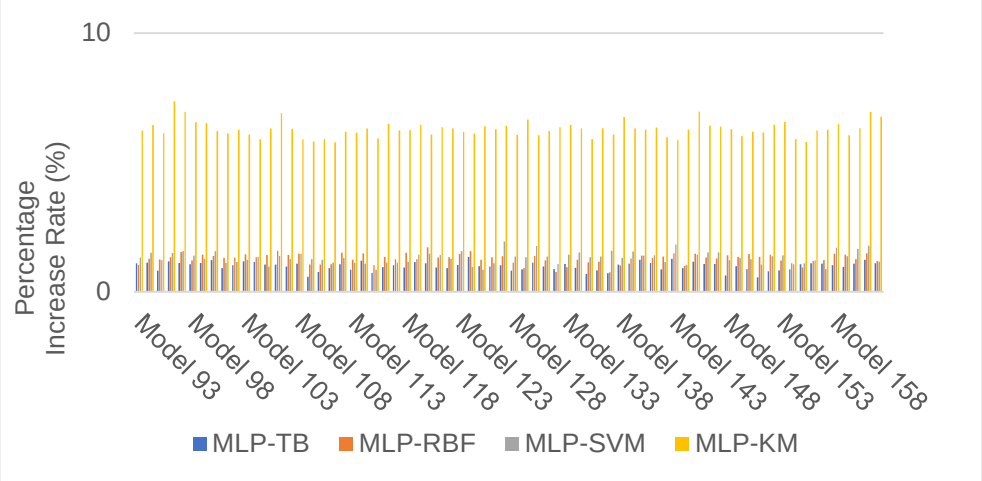


Figure 5.13. Percentage increase rates in accuracy values of COVID-19 prediction for MLP-based models compared to accuracy values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for quadruple combinations

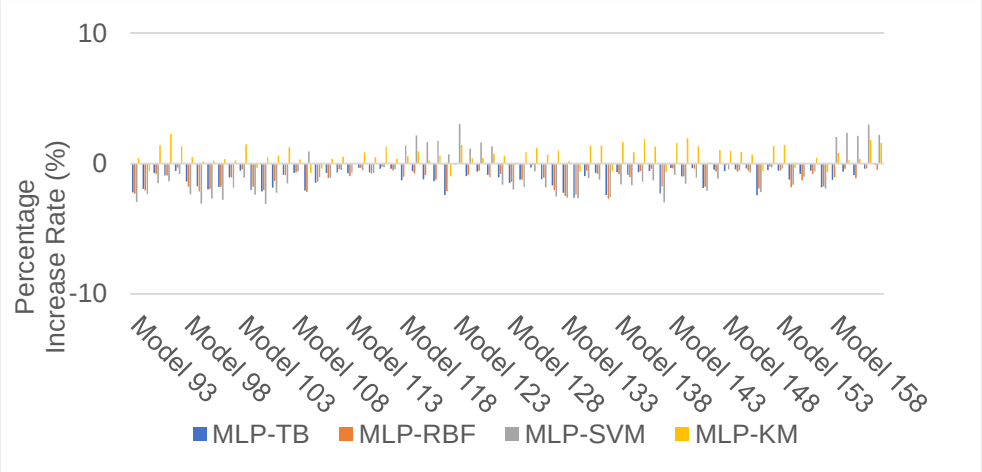


Figure 5.14. Percentage increase rates in precision values of COVID-19 prediction for MLP-based models compared to precision values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for quadruple combinations

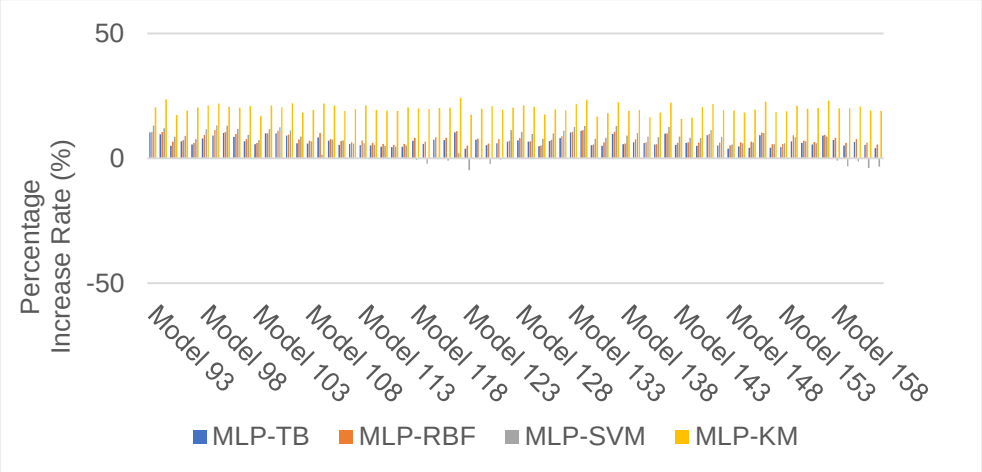


Figure 5.15. Percentage increase rates in recall values of COVID-19 prediction for MLP-based models compared to recall values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for quadruple combinations

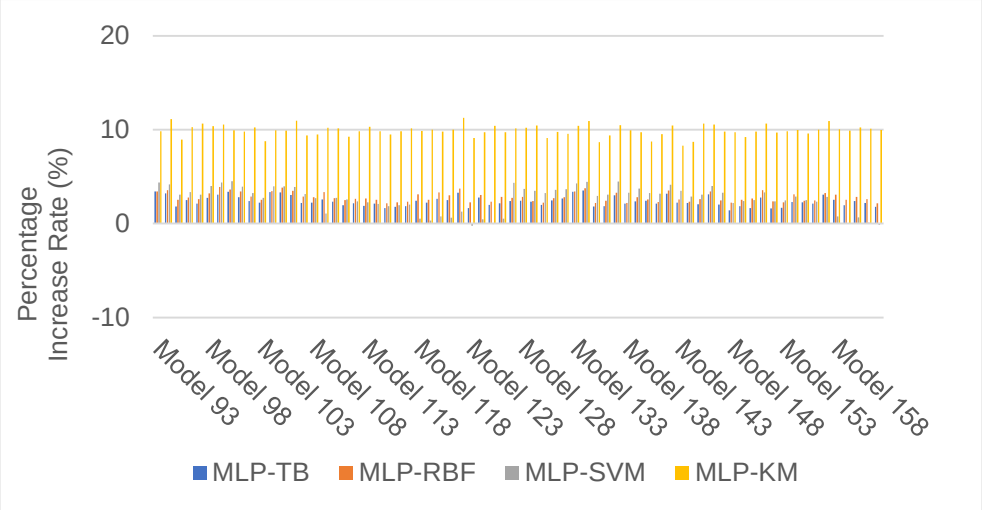


Figure 5.16. Percentage increase rates in F1-score values of COVID-19 prediction for MLP-based models compared to F1-score values obtained by TB-, RBF-,
SVM-, and KMC-based models for quadruple combinations

5.8. Discussion for Prediction Models Using Quadruple Combinations of the Predictor Variables

The following comments regarding the results can be made for the prediction of COVID-19 developed using quadruple combinations of predictor variables:

- When the average accuracy results for each model using all machine learning classifiers are investigated, it is seen that Model 155 including the common variables with lung, renal, diabetes, and obesity comorbidities, exhibits the best performance with an accuracy value of 69.43%. In contrast, Model 106 including the common variables with pregnant, race, diabetes, and obesity comorbidities relatively produces the worst accuracy value of 69.08%.
- When the average precision results for each model using all machine learning classifiers are investigated, it is seen that Model 130 including the common variables with race, lung, cardio, and smoking

comorbidities exhibits the best performance with a precision value of 66.14%. In contrast, Model 109 including the common variables with pregnant, lung, cardio and diabetes comorbidities relatively produces the worst precision value of 65.30%.

- When the average recall results for each model using all machine learning classifiers are investigated, it is seen that Model 109 including the common variables with pregnant, lung, cardio and diabetes comorbidities exhibits the best performance with a recall value of 82.32%. In contrast, Model 130 including the common variables with race, lung, cardio, and smoking comorbidities relatively produces the worst recall value of 78.67%.
- When the average F1-score results for each model using all machine learning classifiers are investigated, it is seen that Model 126 including the common variables with pregnant, renal, smoking, and obesity comorbidities exhibits the best performance with an F1-score value of 72.71%. In contrast, Model 97 including the common variables with pregnant, race, lung and obesity comorbidities relatively produces the worst F1-score value of 71.73%.
- When the average accuracy, recall and F1-score values of models including the quadruple combinations of predictor variables are compared, the general order of machine learning classifiers leading from the best to the worst results is MLP, TB, SVM, RBF, and KMC. When the average precision values of the models including the quadruple combinations of predictor variables are compared, the order of machine learning classifiers leading from the best to worst results is TB, RBF, SVM, MLP, and KMC.
- It is observed that the MLP-based prediction models produce 0.93%, 0.91%, 4.44% and 0.70% higher accuracy values on average compared

to those of SVM-based, RBF-based, K-Means-based, and TB-based prediction models, respectively.

- It is observed that the MLP-based prediction models produce 0.40% higher precision value on average compared to those of K-Means-based prediction models, respectively. In contrast, it is observed that the MLP-based prediction models produce 0.53%, 0.71%, and 0.72% lower precision values compared to those of SVM-based, RBF-based, and TB-based models, respectively.
- It is observed that the MLP-based prediction models produce 5.88%, 6.67%, 17.40% and 5.87% higher recall values on average compared to those of SVM-based, RBF-based, K-Means-based, and TB-based prediction models, respectively.
- It is observed that the MLP-based prediction models produce 1.97%, 2.11%, 7.42%, and 1.79% higher F1-score values on average compared to those of SVM-based, RBF-based, K-Means-based and TB-based prediction models, respectively.
- Average training times of the MLP-based, TB-based, RBF-based, SVM-based, and KMC-based prediction models using computers with 3.20 GHz Dual Core CPU, 3 GB RAM, and 32-bit Windows 7 Professional OS are appr. 21 min, 2 min, 51 min, 9 hr., and 7 min, respectively.

5.9. Results for Prediction Models Using Quintuple Combinations of the Predictor Variables

Table 5.29 through Table 5.36 show the accuracy, precision, recall, and F1-score values for COVID-19 prediction models including the quintuple combinations of the predictor variables, respectively. Figure 5.17 through 5.20 illustrates the percentage increase rates in accuracy, precision, recall, and F1-score values for COVID-19 prediction models.

Table 5.30. Accuracy values for COVID-19 models including the quintuple combinations of the predictor variables – Part 1

Models	Accuracy (%)				
	MLP	TB	RBF	SVM	KMC
Model 163	70.62	69.77	69.81	67.50	66.25
Model 164	70.53	69.91	69.63	69.63	65.91
Model 165	70.67	69.81	69.66	69.47	66.22
Model 166	70.58	69.84	69.55	69.58	66.05
Model 167	70.50	69.92	69.67	69.78	66.20
Model 168	70.51	69.96	69.74	69.70	65.80
Model 169	70.65	69.82	69.73	69.49	66.20
Model 170	70.71	69.97	69.64	69.64	66.01
Model 171	70.69	69.82	69.75	69.56	65.86
Model 172	70.71	69.97	69.91	69.47	65.88
Model 173	70.55	69.84	69.60	69.68	65.84
Model 174	70.56	69.77	69.81	69.53	66.15
Model 175	70.50	69.71	69.61	69.56	65.91
Model 176	70.64	70.03	69.66	69.67	66.21
Model 177	70.64	69.86	69.59	69.66	66.05

Table 5.31. Accuracy values for COVID-19 models including the quintuple combinations of the predictor variables – Part 2

Models	Accuracy (%)				
	MLP	TB	RBF	SVM	KMC
Model 178	70.58	69.89	69.92	69.50	66.58
Model 179	70.67	69.90	70.04	69.81	66.14
Model 180	70.55	69.84	69.64	69.69	65.72
Model 181	70.50	69.92	70.03	69.64	66.19
Model 182	70.57	70.00	69.78	69.64	66.20
Model 183	70.67	69.93	69.77	69.80	66.30
Model 184	70.77	70.09	69.85	69.81	66.73
Model 185	70.61	69.89	69.66	69.75	66.44
Model 186	70.65	70.06	69.76	69.87	66.57
Model 187	70.66	69.92	69.72	69.65	66.03
Model 188	70.52	70.12	69.62	69.68	66.20
Model 189	70.67	70.25	69.66	69.88	66.50
Model 190	70.73	70.03	69.83	69.95	66.50

Table 5.32. Accuracy values for COVID-19 models including the quintuple combinations of the predictor variables – Part 3

Models	Accuracy (%)				
	MLP	TB	RBF	SVM	KMC
Model 191	70.71	70.10	69.95	69.97	66.40
Model 192	70.69	70.09	69.84	70.03	66.26
Model 193	70.73	70.08	69.84	69.50	66.20
Model 194	70.60	70.14	69.73	69.59	66.27
Model 195	70.64	69.84	69.89	69.46	66.42
Model 196	70.70	70.19	69.72	69.44	66.25
Model 197	70.70	70.05	69.56	69.89	66.23
Model 198	70.69	69.77	69.83	69.59	66.17
Model 199	70.75	69.91	69.92	69.41	66.47
Model 200	70.67	69.89	69.78	69.51	66.36
Model 201	70.66	69.95	69.63	69.53	65.96
Model 202	70.62	70.00	69.62	69.59	66.09
Model 203	70.67	69.78	69.91	69.47	66.61
Model 204	70.76	69.97	69.84	69.63	66.30
Model 205	70.55	69.88	69.60	69.62	65.96
Model 206	70.64	69.95	69.78	69.40	66.11
Model 207	70.67	69.89	69.75	69.37	66.33
Model 208	70.67	69.97	69.77	69.57	65.94
Model 209	70.69	69.79	69.71	69.64	66.00
Model 210	70.63	69.88	69.84	69.45	65.99
Model 211	70.53	69.88	69.57	69.56	66.30
Model 212	70.67	70.01	69.70	69.55	66.17
Model 213	70.68	69.99	69.77	69.84	66.37
Model 214	70.63	70.00	69.64	69.68	66.03
Model 215	70.84	70.08	69.81	69.62	66.26
Model 216	70.82	70.12	69.74	69.62	66.20
Model 217	70.69	70.08	69.80	69.97	66.41
Model 218	70.66	69.89	69.69	69.72	66.22

Table 5.33. Precision values for COVID-19 models including the quintuple combinations of the predictor variables – Part 1

Models	Precision (%)				
	MLP	TB	RBF	SVM	KMC
Model 163	65.39	66.23	66.42	63.92	65.66
Model 164	65.24	66.36	66.18	66.83	65.15
Model 165	65.23	66.17	66.26	66.61	65.11
Model 166	65.58	66.26	66.11	66.71	65.13
Model 167	64.77	66.10	66.13	66.75	65.10
Model 168	65.41	66.28	66.19	66.56	64.77
Model 169	65.96	66.01	66.18	66.38	64.94
Model 170	65.16	66.45	66.14	66.67	64.38
Model 171	65.92	66.20	66.28	66.61	64.69
Model 172	65.74	66.38	66.41	66.43	64.76
Model 173	64.93	66.00	66.08	66.81	64.51
Model 174	65.51	66.04	66.22	66.65	64.63
Model 175	65.37	65.88	66.15	66.64	64.34
Model 176	65.33	66.52	66.15	66.84	65.11
Model 177	65.33	66.18	66.12	66.81	65.09
Model 178	65.93	66.32	66.32	66.67	64.86
Model 179	65.57	66.23	66.42	66.78	64.52
Model 180	65.88	66.32	66.08	66.65	64.48
Model 181	65.48	66.36	66.42	66.54	65.30
Model 182	65.63	66.43	66.20	66.70	65.06
Model 183	65.67	66.08	66.27	66.29	65.31
Model 184	65.86	66.13	66.32	66.20	65.68
Model 185	65.90	66.14	66.14	66.26	65.20
Model 186	65.60	66.16	66.27	66.34	65.74
Model 187	65.33	66.11	66.29	66.25	65.08
Model 188	64.91	66.35	66.16	66.18	65.16
Model 189	65.43	66.39	66.14	66.21	65.40
Model 190	65.82	66.27	66.28	66.31	65.52

Table 5.34. Precision values for COVID-19 models including the quintuple combinations of the predictor variables – Part 2

Models	Precision (%)				
	MLP	TB	RBF	SVM	KMC
Model 191	65.40	66.30	66.39	66.26	65.58
Model 192	64.95	66.31	66.34	66.37	65.31
Model 193	65.72	66.30	66.22	64.09	65.11
Model 194	66.07	66.35	66.17	64.16	65.71
Model 195	65.33	66.19	66.32	64.05	65.15
Model 196	65.92	66.45	66.24	63.93	65.14
Model 197	65.87	66.23	66.07	64.49	65.28
Model 198	65.79	66.17	66.29	66.80	65.35
Model 199	65.76	66.31	66.44	66.58	65.50
Model 200	65.60	66.37	66.28	66.66	65.14
Model 201	65.40	66.38	66.19	66.76	64.82
Model 202	65.79	66.23	66.19	66.82	64.96
Model 203	65.75	66.02	66.30	66.66	65.09
Model 204	65.67	66.38	66.36	66.65	65.19
Model 205	65.71	66.24	66.16	66.63	64.87
Model 206	65.84	66.33	66.34	66.37	64.86
Model 207	65.69	66.39	66.28	66.51	64.89
Model 208	65.30	66.39	66.35	66.78	65.54
Model 209	65.95	66.34	66.28	66.82	64.50
Model 210	65.70	66.23	66.41	66.61	64.53
Model 211	65.63	66.30	66.16	66.80	65.23
Model 212	65.76	66.31	66.28	66.62	65.18
Model 213	65.56	66.12	66.29	66.34	65.43
Model 214	65.77	66.25	66.12	66.28	65.09
Model 215	66.24	66.33	66.33	66.15	65.43
Model 216	66.08	66.30	66.25	66.17	65.35
Model 217	65.12	66.21	66.30	66.32	65.65
Model 218	65.99	66.23	66.10	66.30	65.03

Table 5.35. Recall values for COVID-19 models including the quintuple combinations of the predictor variables – Part 1

Models	Recall (%)				
	MLP	TB	RBF	SVM	KMC
Model 163	87.61	80.70	80.14	80.35	68.16
Model 164	87.89	80.75	80.28	77.96	68.40
Model 165	88.50	81.06	80.13	78.08	69.86
Model 166	86.62	80.83	80.23	78.18	69.07
Model 167	89.90	81.76	80.66	78.84	69.84
Model 168	87.06	81.27	80.72	79.19	69.27
Model 169	85.36	81.73	80.71	79.00	70.44
Model 170	89.02	80.65	80.48	78.56	71.68
Model 171	85.69	80.99	80.42	78.45	69.86
Model 172	86.51	80.95	80.55	78.74	69.67
Model 173	89.33	81.81	80.55	78.20	70.45
Model 174	86.84	81.40	80.90	78.20	71.36
Model 175	87.21	81.78	80.33	78.33	71.36
Model 176	87.96	80.66	80.50	78.06	69.86
Model 177	87.96	81.23	80.38	78.15	69.24
Model 178	85.15	80.86	80.98	78.01	72.37
Model 179	87.02	81.20	81.07	78.86	71.73
Model 180	85.25	80.65	80.69	78.81	70.01
Model 181	86.72	80.77	81.04	79.02	69.09
Model 182	86.39	80.90	80.85	78.42	69.97
Model 183	86.66	81.92	80.52	80.57	69.50
Model 184	86.23	82.35	80.68	80.95	70.11
Model 185	85.43	81.52	80.59	80.52	70.51
Model 186	86.84	82.11	80.50	80.66	69.20
Model 187	88.02	81.73	80.26	80.10	69.19
Model 188	89.32	81.66	80.32	80.51	69.64
Model 189	87.65	82.03	80.55	81.22	70.08
Model 190	86.25	81.61	80.72	81.11	69.63

Table 5.36. Recall values for COVID-19 models including the quintuple combinations of the predictor variables – Part 2

Models	Recall (%)				
	MLP	TB	RBF	SVM	KMC
Model 191	87.98	81.75	80.82	81.38	69.01
Model 192	89.86	81.66	80.58	81.20	69.36
Model 193	86.65	81.68	80.99	88.72	69.78
Model 194	84.70	81.75	80.74	88.77	68.04
Model 195	87.96	81.13	80.82	88.73	70.60
Model 196	85.72	81.55	80.45	89.20	69.90
Model 197	85.91	81.83	80.41	88.54	69.36
Model 198	86.22	80.89	80.71	77.92	68.84
Model 199	86.57	80.96	80.52	77.93	69.62
Model 200	86.92	80.63	80.55	78.08	70.39
Model 201	87.75	80.86	80.25	77.78	69.81
Model 202	85.91	81.62	80.21	77.81	69.85
Model 203	86.32	81.53	80.97	77.88	71.62
Model 204	86.99	80.89	80.47	78.56	69.95
Model 205	85.94	81.05	80.23	78.59	69.63
Model 206	85.77	81.01	80.34	78.66	70.34
Model 207	86.53	80.60	80.38	78.04	71.13
Model 208	88.22	80.86	80.21	77.90	67.22
Model 209	85.58	80.36	80.24	78.06	71.19
Model 210	86.33	81.10	80.29	77.97	71.02
Model 211	86.21	80.87	80.12	77.78	69.80
Model 212	86.27	81.36	80.22	78.37	69.42
Model 213	87.13	82.01	80.42	80.53	69.43
Model 214	86.03	81.53	80.56	80.13	69.13
Model 215	85.00	81.57	80.45	80.37	68.96
Model 216	85.57	81.83	80.48	80.28	68.96
Model 217	89.12	82.01	80.55	81.13	68.84
Model 218	85.26	81.18	80.85	80.20	70.15

Table 5.37. F1-score values for COVID-19 models including the quintuple combinations of the predictor variables – Part 1

Models	F1-score (%)				
	MLP	TB	RBF	SVM	KMC
Model 163	74.88	72.75	72.64	71.20	66.89
Model 164	74.89	72.85	72.55	71.97	66.73
Model 165	75.11	72.86	72.54	71.89	67.40
Model 166	74.65	72.82	72.49	71.99	67.04
Model 167	75.30	73.10	72.68	72.29	67.39
Model 168	74.70	73.01	72.73	72.33	66.94
Model 169	74.41	73.03	72.72	72.14	67.58
Model 170	75.25	72.86	72.61	72.13	67.83
Model 171	74.51	72.85	72.67	72.05	67.18
Model 172	74.71	72.94	72.80	72.06	67.12
Model 173	75.20	73.06	72.60	72.06	67.35
Model 174	74.68	72.92	72.83	71.96	67.83
Model 175	74.73	72.97	72.55	72.01	67.67
Model 176	74.97	72.91	72.62	72.02	67.40
Model 177	74.97	72.93	72.55	72.04	67.10
Model 178	74.32	72.87	72.92	71.90	68.41
Model 179	74.79	72.96	73.02	72.32	67.94
Model 180	74.32	72.79	72.66	72.22	67.13
Model 181	74.62	72.86	73.01	72.24	67.14
Model 182	74.59	72.95	72.80	72.09	67.43
Model 183	74.72	73.15	72.70	72.74	67.34
Model 184	74.68	73.36	72.80	72.84	67.82
Model 185	74.40	73.03	72.65	72.69	67.75
Model 186	74.74	73.28	72.69	72.80	67.43
Model 187	75.00	73.09	72.61	72.52	67.07
Model 188	75.19	73.21	72.56	72.64	67.33
Model 189	74.93	73.39	72.64	72.95	67.66
Model 190	74.66	73.14	72.79	72.97	67.51

Table 5.38. F1-score values for COVID-19 models including the quintuple combinations of the predictor variables – Part 2

Models	F1-score (%)				
	MLP	TB	RBF	SVM	KMC
Model 191	75.03	73.22	72.90	73.05	67.25
Model 192	75.40	73.19	72.77	73.04	67.27
Model 193	74.75	73.19	72.86	74.42	67.36
Model 194	74.23	73.25	72.74	74.49	66.85
Model 195	74.97	72.90	72.86	74.39	67.76
Model 196	74.53	73.23	72.65	74.48	67.44
Model 197	74.57	73.21	72.54	74.63	67.26
Model 198	74.63	72.80	72.79	71.93	67.05
Model 199	74.75	72.90	72.80	71.81	67.50
Model 200	74.77	72.81	72.72	71.92	67.66
Model 201	74.94	72.91	72.55	71.85	67.23
Model 202	74.52	73.12	72.53	71.90	67.32
Model 203	74.64	72.96	72.91	71.84	68.20
Model 204	74.84	72.92	72.73	72.12	67.49
Model 205	74.48	72.90	72.52	72.12	67.17
Model 206	74.50	72.94	72.67	71.99	67.49
Model 207	74.68	72.81	72.65	71.81	67.87
Model 208	75.05	72.92	72.62	71.91	66.37
Model 209	74.49	72.68	72.60	72.00	67.68
Model 210	74.61	72.92	72.69	71.85	67.62
Model 211	74.53	72.87	72.47	71.87	67.44
Model 212	74.63	73.07	72.58	72.02	67.23
Model 213	74.82	73.21	72.68	72.75	67.37
Model 214	74.55	73.10	72.63	72.55	67.05
Model 215	74.46	73.16	72.71	72.57	67.15
Model 216	74.57	73.25	72.67	72.55	67.10
Model 217	75.25	73.27	72.73	72.98	67.21
Model 218	74.40	72.95	72.73	72.59	67.49

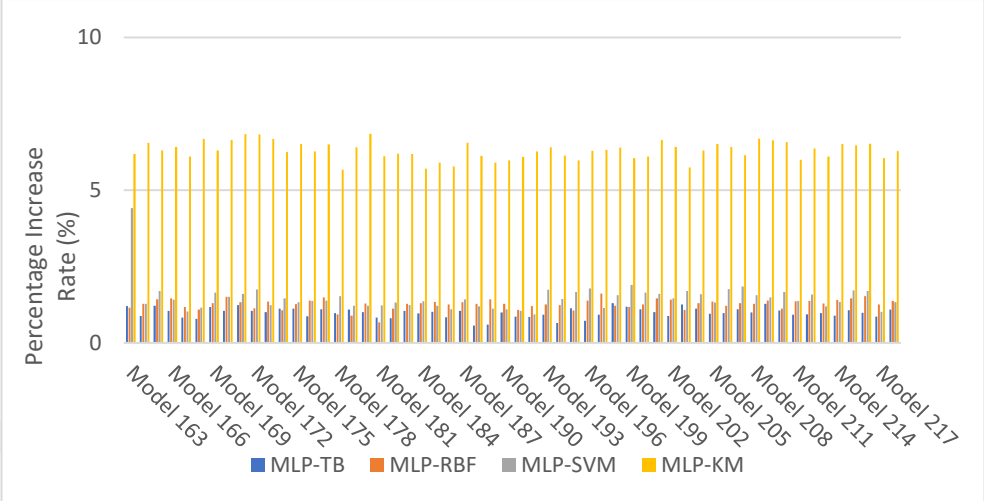


Figure 5.17. Percentage increase rates in accuracy values of COVID-19 prediction for MLP-based models compared to accuracy values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for quintuple combinations

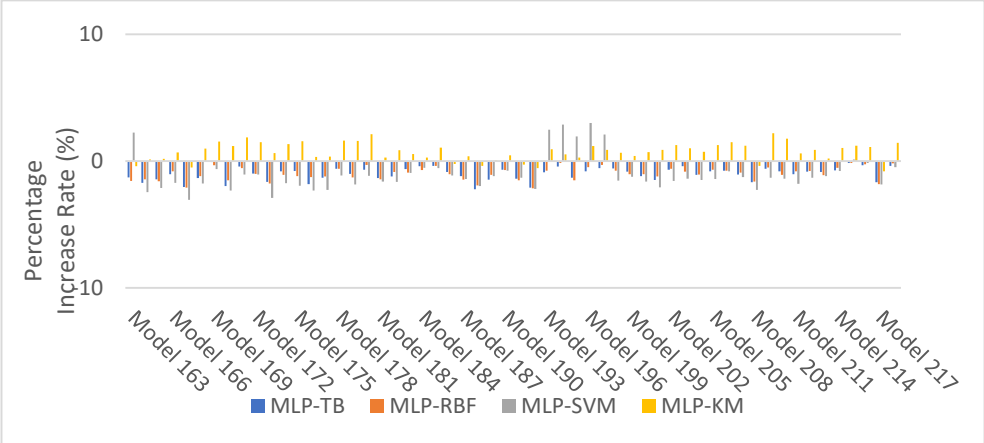


Figure 5.18. Percentage increase rates in precision values of COVID-19 prediction for MLP-based models compared to precision values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for quintuple combinations

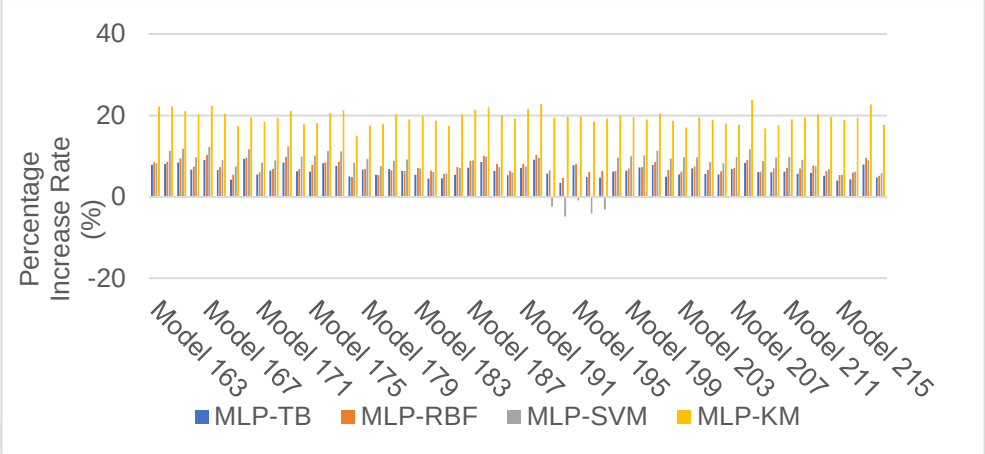


Figure 5.19. Percentage increase rates in recall values of COVID-19 prediction for MLP-based models compared to recall values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for quintuple combinations

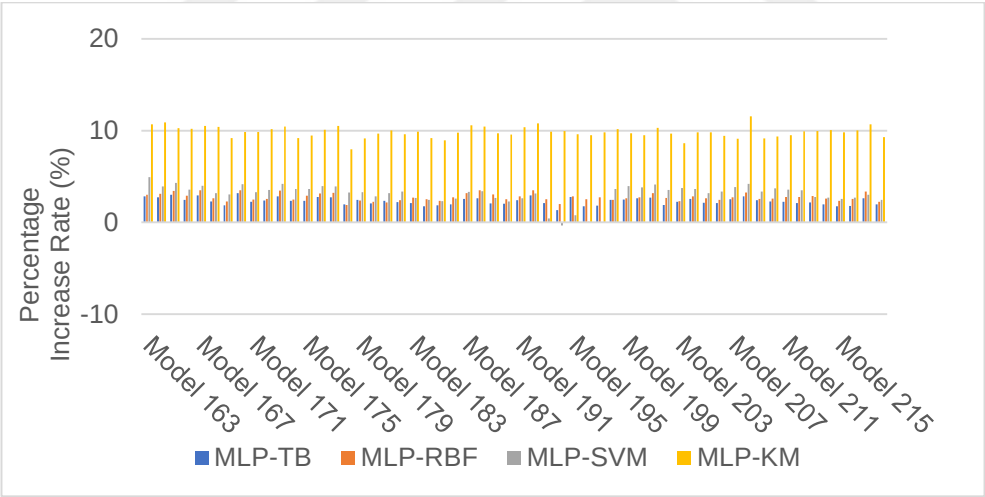


Figure 5.20. Percentage increase rates in F1-score values of COVID-19 prediction for MLP-based models compared to F1-score values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for quintuple combinations

5.10. Discussion for Prediction Models Using Quintuple Combinations of the Predictor Variables

The following comments regarding the results can be made for prediction of COVID-19 developed using quintuple combinations of predictor variables:

- When the average accuracy results for each model using all machine learning classifiers are investigated, it is seen that Model 184 including the common variables with pregnant, lung, cardio, renal, and smoking comorbidities, exhibits the best performance with an accuracy value of 69.45%. In contrast, Model 163 including the common variables with pregnant, race, lung, cardio, and renal comorbidities relatively produces the worst accuracy value of 68.79%.
- When the average precision results for each model using all machine learning classifiers are investigated, it is seen that Model 199 including the common variables with race, lung, cardio, renal. and smoking comorbidities exhibits the best performance with a precision value of 66.12%. In contrast, Model 195 including the common variables with pregnant, cardio, renal, smoking, and obesity comorbidities relatively produces the worst precision value of 65.41%.
- When the average recall results for each model using all machine learning classifiers are investigated, it is seen that Model 195 including the common variables with pregnant, cardio, renal, smoking, and obesity comorbidities exhibits the best performance with a recall value of 81.85%. In contrast, Model 208 including the common variables with race, cardio, renal, diabetes, and smoking comorbidities relatively produces the worst recall value of 78.88%.
- When the average F1-score results for each model using all machine learning classifiers are investigated, it is seen that Model 195 including the common variables with pregnant, cardio, renal, smoking, and

obesity comorbidities exhibits the best performance with an F1-score value of 72.58%. In contrast, Model 163 including the common variables with pregnant, race, lung, cardio, and renal comorbidities relatively produces the worst F1-score value of 71.67%.

- When the average accuracy, recall and F1-score values of models including the quintuple combinations of predictor variables are compared, the general order of machine learning classifiers leading from the best to the worst results is MLP, TB, RBF, SVM, and KMC. When the average precision values of the models including the quintuple combinations of predictor variables are compared, the order of machine learning classifiers leading from the best to worst results is SVM, TB, RBF, MLP, and KMC.
- It is observed that the MLP-based prediction models produce 1.04%, 0.90%, 4.45% and 0.70% higher accuracy values on average compared to those of SVM-based, RBF-based, K-Means-based, and TB-based prediction models, respectively.
- It is observed that the MLP-based prediction models produce 0.49% higher precision value on average compared to those of K-Means-based prediction models, respectively. In contrast, it is observed that the MLP-based prediction models produce 0.68%, 0.65%, and 0.66% lower precision values compared to those of SVM-based, RBF-based, and TB-based models, respectively.
- It is observed that the MLP-based prediction models produce 6.92%, 6.34%, 17.03% and 5.57% higher recall values on average compared to those of SVM-based, RBF-based, K-Means-based, and TB-based prediction models, respectively.
- It is observed that the MLP-based prediction models produce 2.30%, 2.04%, 7.36%, and 1.73% higher F1-score values on average compared

to those of SVM-based, RBF-based, K-Means-based and TB-based prediction models, respectively.

- Average training times of the MLP-based, TB-based, RBF-based, SVM-based, and KMC-based prediction models using computers with 3.20 GHz Dual Core CPU, 3 GB RAM, and 32-bit Windows 7 Professional OS are appr. 23 min, 2 min, 59 min, 8 hr., and 7 min, respectively.

5.11. Results for Prediction Models Using Sextuple Combinations of the Predictor Variables

Table 5.37 through Table 5.40 show the accuracy, precision, recall, and F1-score values for COVID-19 prediction models including the sextuple combinations of the predictor variables, respectively. Figure 5.21 through 5.24 illustrates the percentage increase rates in accuracy, precision, recall, and F1-score values for COVID-19 prediction models.

Table 5.39. Accuracy values for COVID-19 models including the sextuple combinations of the predictor variables – Part 1

Models	Accuracy (%)				
	MLP	TB	RBF	SVM	KMC
Model 219	70.66	69.84	69.67	69.56	66.25
Model 220	70.81	69.80	69.78	69.48	66.15
Model 221	70.53	69.85	69.69	69.49	66.00
Model 222	70.70	69.97	69.64	69.52	65.95
Model 223	70.59	70.02	69.73	69.58	66.00
Model 224	70.69	69.89	69.83	69.46	66.14
Model 225	70.72	70.02	69.63	69.60	66.09
Model 226	70.48	69.87	69.76	69.66	67.71
Model 227	70.59	69.95	69.94	69.46	65.85
Model 228	70.59	70.00	69.95	69.42	65.98
Model 229	70.88	69.92	69.88	69.58	66.09
Model 230	70.67	69.78	69.69	69.64	65.81
Model 231	70.49	69.89	70.03	69.41	66.14
Model 232	70.53	69.84	69.79	69.51	66.28
Model 233	70.63	69.89	69.70	69.53	66.22

Table 5.40. Accuracy values for COVID-19 models including the sextuple combinations of the predictor variables – Part 2

Models	Accuracy (%)				
	MLP	TB	RBF	SVM	KMC
Model 234	70.75	69.98	69.94	69.77	66.52
Model 235	70.68	69.99	69.70	69.66	65.96
Model 236	70.64	70.07	69.95	69.64	66.39
Model 237	70.70	69.99	69.84	69.63	66.20
Model 238	70.70	70.09	69.88	69.96	66.36
Model 239	70.64	70.19	69.66	69.30	66.24
Model 240	70.70	70.03	69.70	69.45	66.20
Model 241	70.67	69.95	69.88	69.66	66.04
Model 242	70.49	69.93	69.79	69.43	66.47
Model 243	70.69	69.98	69.84	69.39	66.33
Model 244	70.80	69.89	69.81	69.30	66.05
Model 245	70.65	69.98	69.65	69.59	65.93
Model 246	70.88	70.10	69.73	69.62	66.33

Table 5.41. Precision values for COVID-19 models including the sextuple combinations of the predictor variables

Models	Precision (%)				
	MLP	TB	RBF	SVM	KMC
Model 219	65.93	66.29	66.24	66.82	65.37
Model 220	65.96	66.30	66.30	66.66	65.21
Model 221	65.40	66.27	66.33	66.64	64.50
Model 222	65.49	66.41	66.24	66.81	64.71
Model 223	65.88	66.40	66.31	66.86	64.80
Model 224	65.93	66.23	66.32	66.67	64.42
Model 225	65.97	66.23	66.21	66.65	64.41
Model 226	65.83	66.25	66.32	66.71	65.57
Model 227	65.66	66.29	66.43	66.44	65.33
Model 228	65.92	66.26	66.56	66.59	64.88
Model 229	65.98	66.13	66.43	66.81	65.46
Model 230	65.76	65.97	66.21	66.82	64.53
Model 231	64.71	66.30	66.49	66.59	64.69
Model 232	65.36	66.24	66.37	66.76	65.30
Model 233	65.80	66.15	66.25	66.62	65.50
Model 234	65.77	66.04	66.30	66.31	65.69
Model 235	66.03	66.24	66.15	66.29	64.87
Model 236	65.52	66.18	66.34	66.20	65.51
Model 237	65.91	66.19	66.27	66.22	65.11
Model 238	65.75	66.26	66.29	66.34	65.42
Model 239	66.00	66.43	66.12	63.80	65.17
Model 240	65.44	66.46	66.21	66.74	65.42
Model 241	65.36	66.25	66.39	66.90	65.07
Model 242	65.15	66.57	66.31	66.65	64.83
Model 243	65.94	66.40	66.38	66.74	64.80
Model 244	65.88	66.35	66.37	66.46	64.62
Model 245	65.81	66.49	66.22	66.82	65.19
Model 246	65.99	66.07	66.15	66.23	65.60

Table 5.42. Recall values for COVID-19 models including the sextuple combinations of the predictor variables

Models	Recall (%)				
	MLP	TB	RBF	SVM	KMC
Model 219	85.50	80.72	80.25	77.68	69.10
Model 220	86.03	80.55	80.48	77.93	69.24
Model 221	87.19	80.86	79.97	78.05	71.18
Model 222	87.53	80.84	80.09	77.56	70.18
Model 223	85.46	81.04	80.22	77.64	70.04
Model 224	85.63	81.18	80.60	77.84	72.14
Model 225	85.58	81.69	80.17	78.46	71.90
Model 226	85.18	81.02	80.31	78.50	74.59
Model 227	86.36	81.21	80.62	78.63	67.55
Model 228	85.26	81.49	80.15	77.95	69.68
Model 229	86.20	81.66	80.36	77.85	68.14
Model 230	86.21	81.71	80.44	78.02	70.19
Model 231	90.14	80.93	80.74	77.89	71.09
Model 232	87.37	80.91	80.25	77.71	69.47
Model 233	85.90	81.50	80.32	78.26	68.53
Model 234	86.58	82.26	81.10	80.39	69.16
Model 235	85.17	81.56	80.70	79.97	69.63
Model 236	87.15	82.08	81.02	80.27	69.21
Model 237	85.78	81.74	80.78	80.13	69.79
Model 238	86.44	81.87	80.91	81.03	69.44
Model 239	85.12	81.66	80.65	89.23	69.78
Model 240	87.77	80.91	80.46	77.55	68.72
Model 241	87.96	81.35	80.52	77.80	69.27
Model 242	88.15	80.06	80.47	77.77	71.97
Model 243	85.60	80.89	80.42	77.31	71.52
Model 244	86.29	80.71	80.34	77.94	70.90
Model 245	85.97	80.60	80.23	77.82	68.37
Model 246	86.21	82.63	80.82	80.07	68.67

Table 5.43. F1-score values for COVID-19 models including the sextuple combinations of the predictor variables

Models	F1-score (%)				
	MLP	**TB**	**RBF**	**SVM**	**KMC**
Model 219	74.45	72.80	72.58	71.84	67.19
Model 220	74.67	72.73	72.70	71.86	67.16
Model 221	74.74	72.84	72.52	71.90	67.68
Model 222	74.92	72.92	72.51	71.78	67.34
Model 223	74.40	72.99	72.60	71.85	67.32
Model 224	74.50	72.95	72.76	71.82	68.06
Model 225	74.51	73.15	72.53	72.07	67.95
Model 226	74.26	72.89	72.65	72.12	69.79
Model 227	74.60	72.99	72.84	72.03	66.42
Model 228	74.35	73.09	72.73	71.82	67.20
Model 229	74.75	73.08	72.73	71.91	66.77
Model 230	74.61	73.00	72.63	71.99	67.24
Model 231	75.34	72.89	72.93	71.80	67.74
Model 232	74.78	72.85	72.65	71.82	67.32
Model 233	74.52	73.03	72.61	71.97	66.98
Model 234	74.75	73.26	72.96	72.67	67.38
Model 235	74.39	73.11	72.70	72.49	67.17
Model 236	74.80	73.28	72.95	72.56	67.31
Model 237	74.54	73.15	72.81	72.52	67.37
Model 238	74.69	73.25	72.88	72.95	67.37
Model 239	74.35	73.26	72.67	74.40	67.39
Model 240	74.98	72.97	72.64	71.74	67.03
Model 241	74.99	73.03	72.78	71.94	67.10
Model 242	74.92	72.70	72.71	71.78	68.21
Model 243	74.49	72.93	72.73	71.64	67.99
Model 244	74.72	72.83	72.69	71.74	67.62
Model 245	74.55	72.87	72.55	71.90	66.74
Model 246	74.75	73.43	72.75	72.49	67.10

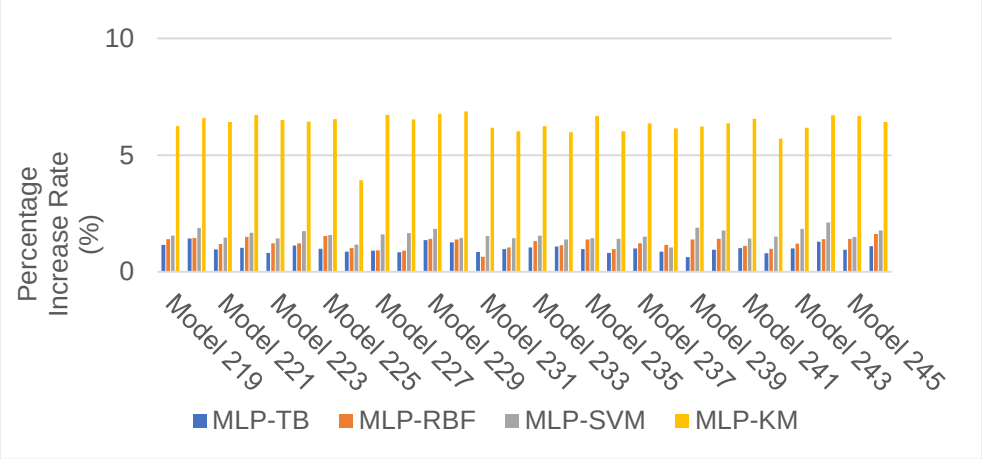


Figure 5.21. Percentage increase rates in accuracy values of COVID-19 prediction for MLP-based models compared to accuracy values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for sextuple combinations

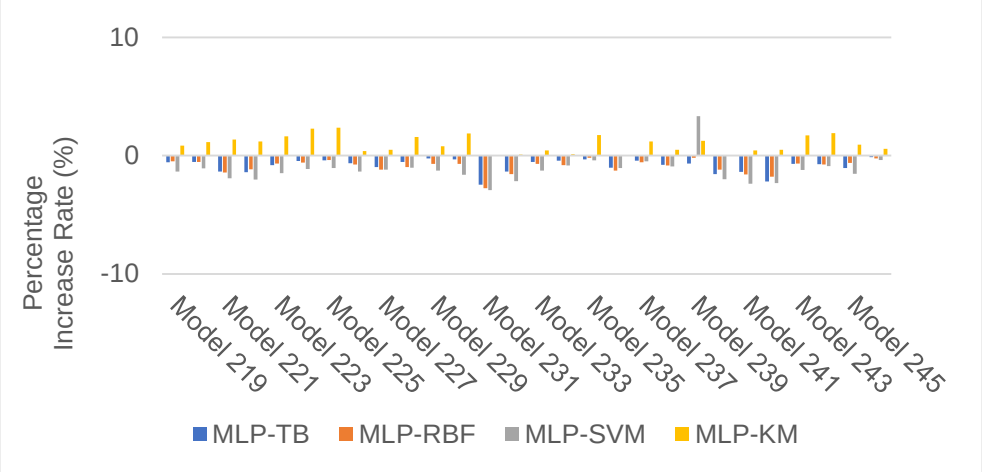


Figure 5.22. Percentage increase rates in precision values of COVID-19 prediction for MLP-based models compared to precision values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for sextuple combinations

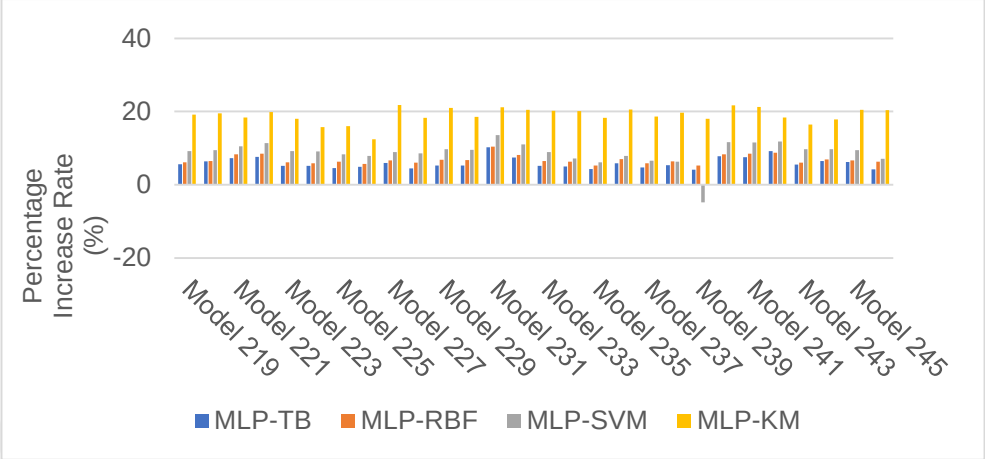


Figure 5.23. Percentage increase rates in recall values of COVID-19 prediction for MLP-based models compared to recall values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for sextuple combinations

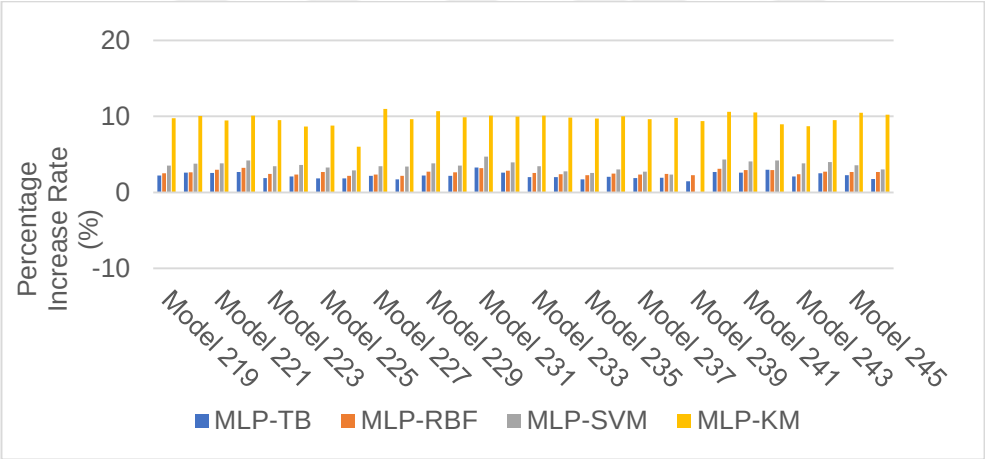


Figure 5.24. Percentage increase rates in F1-score values of COVID-19 prediction for MLP-based models compared to F1-score values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for sextuple combinations

5.12. Discussion for Prediction Models Using Sextuple Combinations of the Predictor Variables

The following comments regarding the results can be made for prediction of COVID-19 developed using sextuple combinations of predictor variables:

- When the average accuracy results for each model using all machine learning classifiers are investigated, it is seen that Model 226 including the common variables with pregnant, race, lung, renal, diabetes and obesity comorbidities, exhibits the best performance with an accuracy value of 69.50%. In contrast, Model 221 including the common variables with pregnant, race, lung, cardio, renal, and obesity comorbidities relatively produces the worst accuracy value of 69.11%.
- When the average precision results for each model using all machine learning classifiers are investigated, it is seen that Model 229 including the common variables with pregnant, race, cardio, renal, diabetes, and smoking comorbidities exhibits the best performance with a precision value of 66.16%. In contrast, Model 239 including the common variables with pregnant, cardio, renal, diabetes, smoking, and obesity comorbidities relatively produces the worst precision value of 65.50%.
- When the average recall results for each model using all machine learning classifiers are investigated, it is seen that Model 239 including the common variables with pregnant, cardio, renal, diabetes, smoking, and obesity comorbidities exhibits the best performance with a recall value of 81.29%. In contrast, Model 245 including the common variables with race, cardio, renal, diabetes, smoking, and obesity comorbidities relatively produces the worst recall value of 78.60%.
- When the average F1-score results for each model using all machine learning classifiers are investigated, it is seen that Model 239 including the common variables with pregnant, cardio, renal, diabetes, smoking,

and obesity comorbidities exhibits the best performance with an F1-score value of 72.41%. In contrast, Model 245 including the common variables with race, cardio, renal, diabetes, smoking, and obesity comorbidities relatively produces the worst F1-score value of 71.72%.

- When the average accuracy, recall and F1-score values of models including the sextuple combinations of predictor variables are compared, the general order of machine learning classifiers leading from the best to the worst results is MLP, TB, RBF, SVM, and KMC. When the average precision values of the models including the sextuple combinations of predictor variables are compared, the order of machine learning classifiers leading from the best to worst results is SVM, RBF, TB, MLP, and KMC.
- It is observed that the MLP-based prediction models produce 1.12%, 0.87%, 4.46% and 0.71% higher accuracy values on average compared to those of SVM-based, RBF-based, K-Means-based, and TB-based prediction models, respectively.
- It is observed that the MLP-based prediction models produce 0.65% higher precision values on average compared to those of K-Means-based prediction models, respectively. In contrast, it is observed that the MLP-based prediction models produce 0.79%, 0.59% and 0.55% lower precision values compared to those of SVM-based, RBF-based and TB-based models, respectively.
- It is observed that the MLP-based prediction models produce 7.59%, 5.94%, 16.44% and 5.15% higher recall values on average compared to those of SVM-based, RBF-based, K-Means-based, and TB-based prediction models, respectively.
- It is observed that the MLP-based prediction models produce 2.53%, 1.95%, 7.23%, and 1.64% higher F1-score values on average compared

to those of SVM-based, RBF-based, K-Means-based and TB-based prediction models, respectively.

- Average training times of the MLP-based, TB-based, RBF-based, SVM-based, and KMC-based prediction models using computers with 3.20 GHz Dual Core CPU, 3 GB RAM, and 32-bit Windows 7 Professional OS are appr. 26 min, 2 min, 1 hr., 8 hr., and 8 min, respectively.

5.13. Results for Prediction Models Using Septuple Combinations of the Predictor Variables

Table 5.41 through Table 5.44 show the accuracy, precision, recall, and F1-score values for COVID-19 prediction models including the unique combinations of the predictor variables, respectively. Figure 5.25 through 5.28 illustrates the percentage increase rates in accuracy, precision, recall, and F1-score values for COVID-19 prediction models.

Table 5.44 Accuracy values for COVID-19 models including the septuple combinations of the predictor variables

Models	Accuracy (%)				
	MLP	TB	RBF	SVM	KMC
Model 247	70.71	70.19	69.91	69.45	66.11
Model 248	70.46	69.97	69.74	69.58	65.98
Model 249	70.60	69.97	69.85	69.41	66.01
Model 250	70.62	69.78	69.88	69.38	66.06
Model 251	70.56	70.04	69.72	69.38	65.81
Model 252	70.58	69.84	69.84	69.49	66.05
Model 253	70.78	70.08	69.72	69.58	66.44
Model 254	70.67	69.82	69.71	69.42	66.30

Table 5.45 Precision values for COVID-19 models including the septuple combinations of the predictor variables

Models	Precision (%)				
	MLP	TB	RBF	SVM	KMC
Model 247	65.34	66.63	66.40	66.80	65.36
Model 248	64.98	66.65	66.24	66.87	64.96
Model 249	65.61	66.42	66.38	66.66	64.41
Model 250	65.30	66.24	66.42	66.76	64.73
Model 251	64.74	66.27	66.25	66.56	64.86
Model 252	64.86	66.35	66.38	66.76	65.09
Model 253	66.00	66.25	66.15	66.23	65.66
Model 254	65.67	66.05	66.33	66.79	65.18

Table 5.46 Recall values for COVID-19 models including the septuple combinations of the predictor variables

Models	Recall (%)				
	MLP	TB	RBF	SVM	KMC
Model 247	88.23	80.90	80.58	77.36	68.53
Model 248	88.73	79.93	80.51	77.58	69.40
Model 249	86.60	80.77	80.44	77.63	71.58
Model 250	87.99	80.69	80.40	77.23	70.58
Model 251	90.29	81.61	80.38	77.88	68.97
Model 252	89.86	80.53	80.39	77.63	69.21
Model 253	85.69	81.86	80.76	79.87	68.92
Model 254	86.61	81.57	80.05	77.25	69.99

Table 5.47 F1-score values for COVID-19 models including the septuple combinations of the predictor variables

Models	F1-score (%)				
	MLP	TB	RBF	SVM	KMC
Model 247	75.08	73.08	72.81	71.69	66.91
Model 248	75.02	72.69	72.68	71.83	67.11
Model 249	74.66	72.89	72.74	71.73	67.80
Model 250	74.97	72.75	72.74	71.61	67.53
Model 251	75.41	73.15	72.63	71.78	66.85
Model 252	75.34	72.76	72.72	71.79	67.09
Model 253	74.57	73.23	72.73	72.41	67.25
Model 254	74.70	72.99	72.55	71.64	67.50

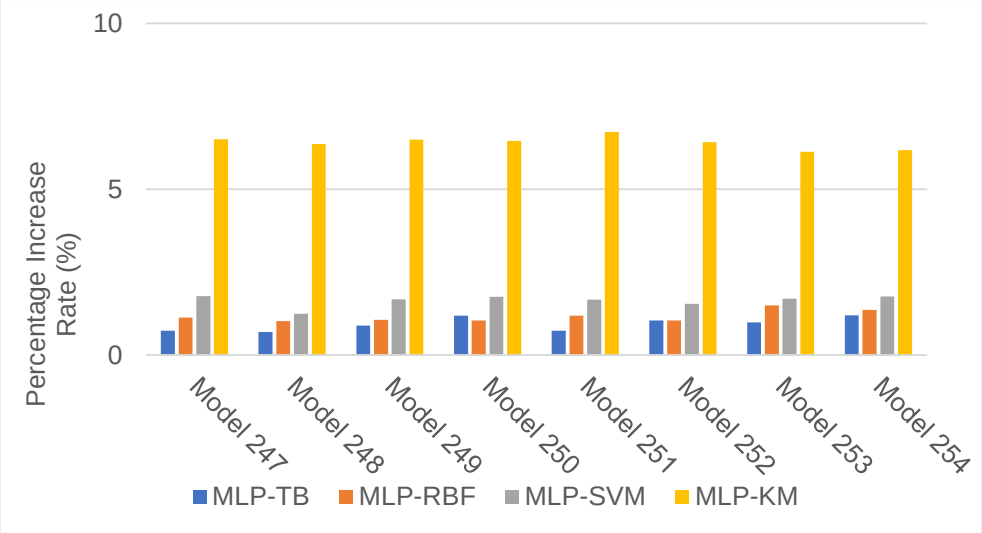


Figure 5.25. Percentage increase rates in accuracy values of COVID-19 prediction for MLP-based models compared to accuracy values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for septuple combinations

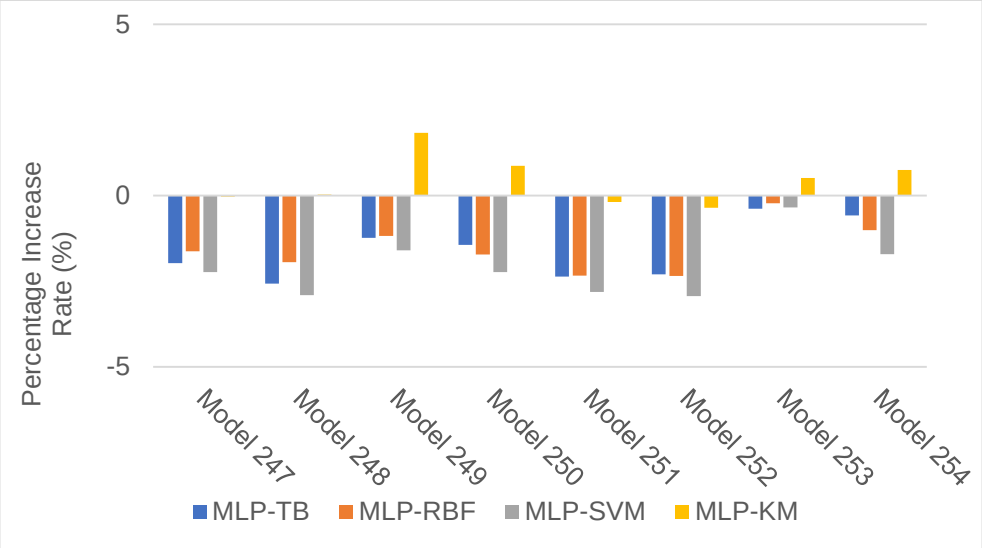


Figure 5.26. Percentage increase rates in precision values of COVID-19 prediction for MLP-based models compared to precision values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for septuple combinations

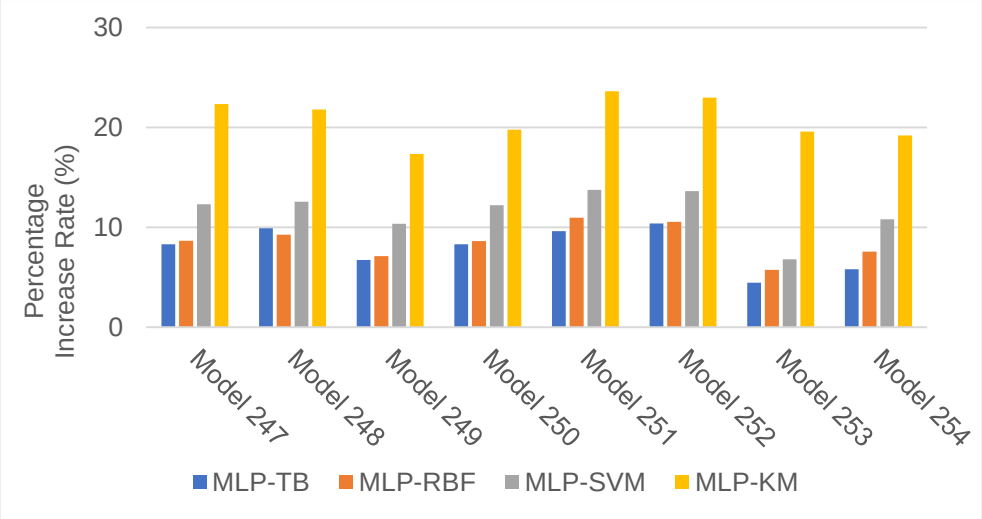


Figure 5.27. Percentage increase rates in recall values of COVID-19 prediction for MLP-based models compared to recall values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for septuple combinations

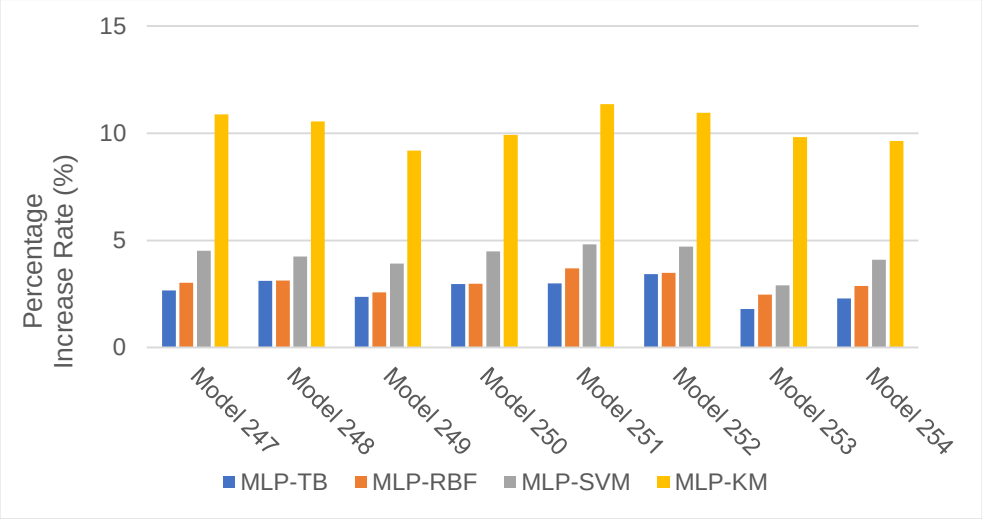


Figure 5.28. Percentage increase rates in F1-score values of COVID-19 prediction for MLP-based models compared to F1-score values obtained by TB-based, RBF-based, SVM-based, and KMC-based models for septuple combinations

5.14. Discussion for Prediction Models Using Septuple Combinations of the Predictor Variables

The following comments regarding the results can be made for the prediction of COVID-19 developed using septuple combinations of predictor variables:

- When the average accuracy results for each model using all machine learning classifiers are investigated, it is seen that Model 253 including the common variables with pregnant, lung, cardio, renal, diabetes, smoking, and obesity comorbidities, exhibits the best performance with an accuracy value of 69.32%. In contrast, Model 251 including the common variables with pregnant, race, lung, renal, diabetes, smoking, and obesity comorbidities relatively produces the worst accuracy value of 69.10%.
- When the average precision results for each model using all machine learning classifiers are investigated, it is seen that Model 247 including the common variables with pregnant, race, lung, cardio, renal, diabetes, and smoking comorbidities exhibits the best performance with a precision value of 66.11%. In contrast, Model 251 including the common variables with pregnant, race, lung, renal, diabetes, smoking, and obesity comorbidities relatively produces the worst precision value of 65.74%.
- When the average recall results for each model using all machine learning classifiers are investigated, it is seen that Model 251 including the common variables with pregnant, race, lung, renal, diabetes, smoking, and obesity comorbidities exhibits the best performance with a recall value of 79.83%. In contrast, Model 254 including the common variables with race, lung, cardio, renal, diabetes, smoking, and obesity comorbidities relatively produces the worst recall value of 79.09%.

- When the average F1-score results for each model using all machine learning classifiers are investigated, it is seen that Model 253 including the common variables with pregnant, lung, cardio, renal, diabetes, smoking, and obesity comorbidities exhibits the best performance with an F1-score value of 72.04%. In contrast, Model 248 including the common variables with pregnant, race, lung, cardio, renal, diabetes, and obesity comorbidities relatively produces the worst F1-score value of 71.87%.
- When the average accuracy, recall and F1-score values of models including the septuple combinations of predictor variables are compared, the general order of machine learning classifiers leading from the best to the worst results is MLP, TB, RBF, SVM, and KMC. When the average precision values of the models including the septuple combinations of predictor variables are compared, the order of machine learning classifiers leading from the best to worst results is SVM, TB, RBF, MLP, and KMC.
- It is observed that the MLP-based prediction models produce 1.16%, 0.83%, 4.53% and 0.66% higher accuracy values on average compared to those of SVM-based, RBF-based, K-Means-based, and TB-based prediction models, respectively.
- It is observed that the MLP-based prediction models produce 0.28% higher precision values on average compared to those of K-Means-based prediction models. In contrast, it is observed that the MLP-based prediction models produce 1.37%, 1.01%, and 1.05% lower precision values compared to those of SVM-based, RBF-based and TB-based models, respectively.
- It is observed that the MLP-based prediction models produce 10.20%, 7.56%, 18.35%, and 7.02% higher recall values on average compared

to those of SVM-based, RBF-based, K-Means-based, and TB-based prediction models, respectively.

- It is observed that the MLP-based prediction models produce 3.16%, 2.27%, 7.71%, and 2.03% higher F1-score values on average compared to those of SVM-based, RBF-based, K-Means-based and TB-based prediction models, respectively.
- Average training times of the MLP-based, TB-based, RBF-based, SVM-based, and KMC-based prediction models using computers with 3.20 GHz Dual Core CPU, 3 GB RAM, and 32-bit Windows 7 Professional OS are appr. 18 min, 2 min, 1 hr., 8 hr., and 8 min, respectively.

5.15. Results for Prediction Models Using Octuple Combinations of the Predictor Variables

Table 5.45 show the accuracy, precision, recall, and F1-score values for COVID-19 prediction models including the octuple combinations of the predictor variables. Figure 5.29 through 5.32 illustrates the percentage increase rates in accuracy, precision, recall, and F1-score values for COVID-19 prediction models.

Table 5.48 Metric values for Model 255 of octuple combination for COVID-19 prediction models.

Metric	MLP	TB	RBF	SVM	KMC
Accuracy (%)	70.64	69.89	69.88	68.46	66.20
Precision (%)	65.33	66.23	66.43	65.37	64.94
Recall (%)	87.96	81.18	80.36	78.53	70.44
F1-score (%)	74.97	72.95	72.73	71.35	67.58

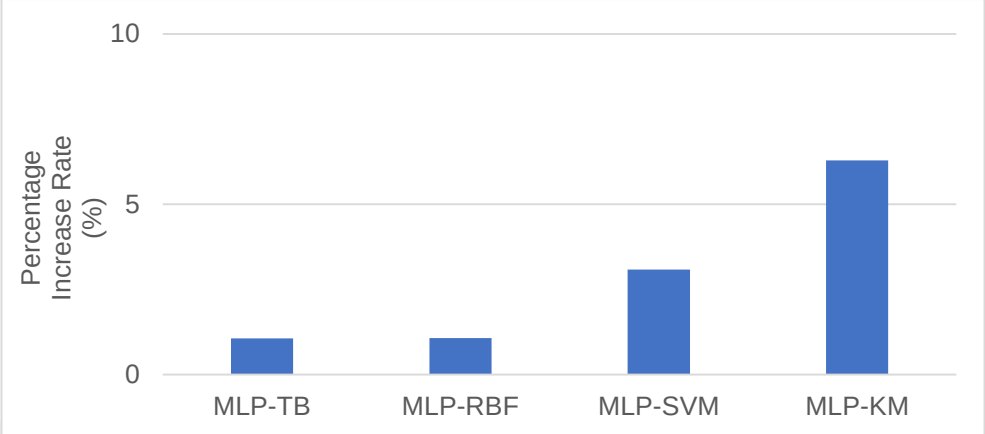


Figure 5.29. Percentage increase rates in accuracy values of COVID-19 prediction for the MLP-based Model 255 compared to accuracy values obtained by TB-based, RBF-based, SVM-based, and KMC-based Model 255 for the octuple combination

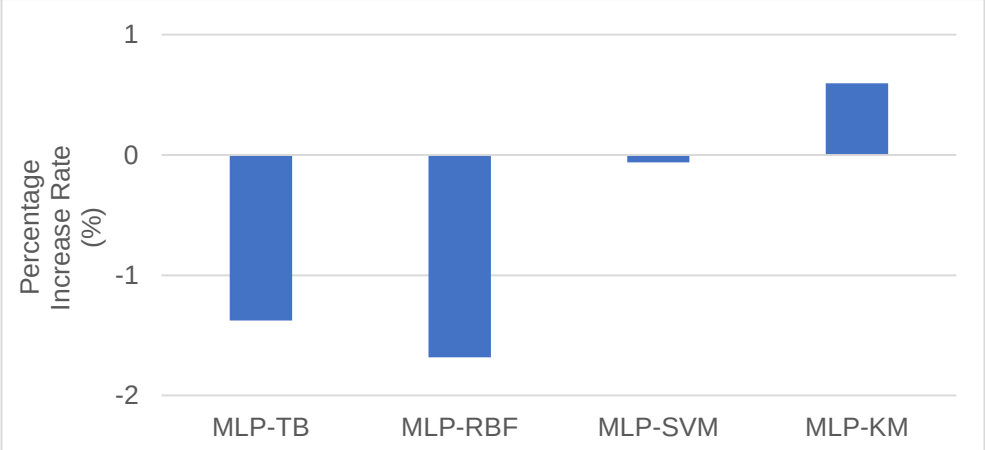


Figure 5.30. Percentage increase rates in precision values of COVID-19 prediction for the MLP-based Model 255 compared to precision values obtained by TB-based, RBF-based, SVM-based, and KMC-based Model 255 for the octuple combination

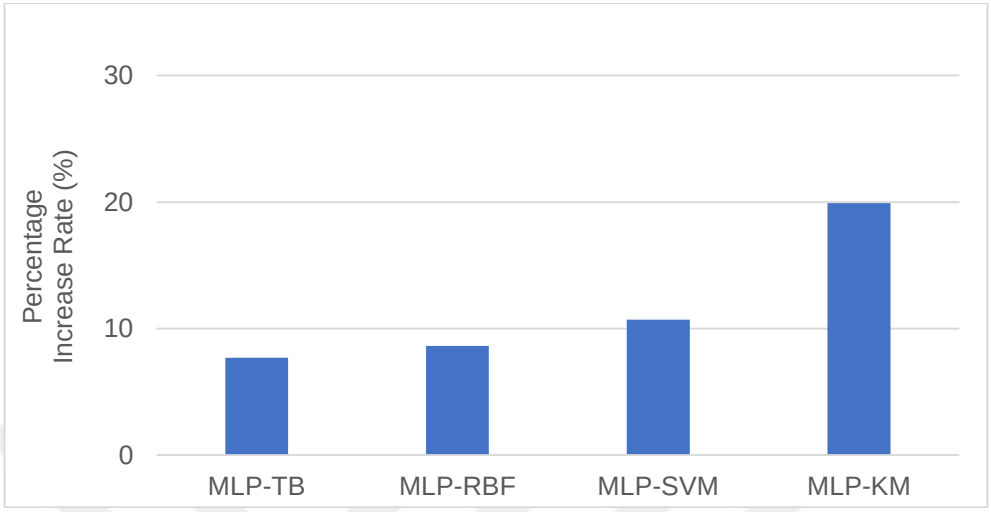


Figure 5.31. Percentage increase rates in recall values of COVID-19 prediction for the MLP-based Model 255 compared to recall values obtained by TB-based, RBF-based, SVM-based, and KMC-based Model 255 the octuple combination

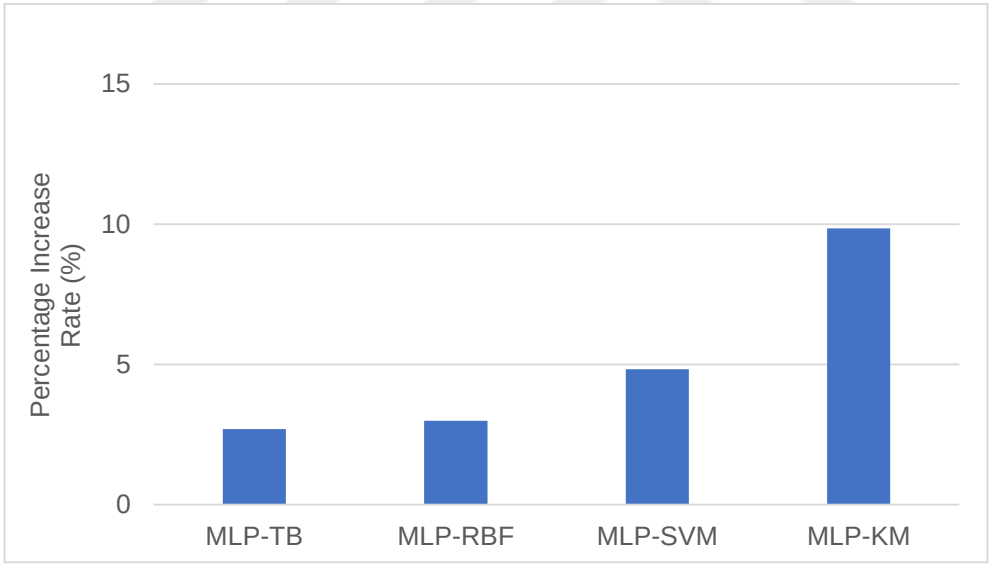


Figure 5.32. Percentage increase rates in F1-score values of COVID-19 prediction for the MLP-based Model 255 compared to F1-score values obtained by TB-based, RBF-based, SVM-based, and KMC-based Model 255 for the octuple combination

5.16. Discussion for Prediction Models Using the Octuple Combination of the Predictor Variables

The following comments regarding the results can be made for the prediction of COVID-19 developed using the octuple combination of the predictor variables:

- When the accuracy, recall and F1-score values of models including the octuple combination of predictor variables are compared, the general order of machine learning classifiers leading from the best to the worst results is MLP, TB, RBF, SVM, and KMC. When the precision values of the models including the octuple combination of predictor variables are compared, the order of machine learning classifiers leading from the best to worst results is RBF, TB, SVM, MLP, and KMC.
- It is observed that the MLP-based prediction model produces 2.18%, 0.76%, 4.44% and 0.75% higher accuracy value compared to those of SVM-based, RBF-based, K-Means-based, and TB-based prediction models, respectively.
- It is observed that the MLP-based prediction model produces 0.39% higher precision value compared to those of K-Means-based prediction model. In contrast, it is observed that the MLP-based prediction models produce 0.04%, 1.10%, and 0.90% lower precision values compared to those of SVM-based, RBF-based, and TB-based models, respectively.
- It is observed that the MLP-based prediction model produces 9.43%, 7.60%, 17.52%, and 6.78% higher recall values compared to those of SVM-based, RBF-based, K-Means-based, and TB-based prediction models, respectively.
- It is observed that the MLP-based prediction model produces 3.62%, 2.24%, 7.39%, and 2.02% higher F1-score values compared to those of SVM-based, RBF-based, K-Means-based and TB-based prediction models, respectively.

- Average training times of the MLP-based, TB-based, RBF-based, SVM-based, and KMC-based prediction models using computers with 3.20 GHz Dual Core CPU, 3 GB RAM, and 32-bit Windows 7 Professional OS are appr. 18 min, 2 min, 1 hr., 55 min, and 4 min, respectively.

5.17. Results for Feature Selection-Based Prediction Models

5.17.1. Relief-F Results

Table 5.46 through Table 5.49 show the accuracy, precision, recall, and F1-score values for COVID-19 prediction models including the Relief-F combinations of the predictor variables, respectively. Figure 5.33 through 5.36 illustrates the percentage increase rates in accuracy, precision, recall, and F1-score values for COVID-19 prediction models.

Table 5.49. Accuracy values for Relief-F-based COVID-19 prediction models using different machine learning methods

Models	Accuracy (%)				
	MLP	TB	RBF	SVM	KMC
Model 256	70.64	69.89	69.88	69.38	66.20
Model 257	70.62	69.78	69.88	69.38	66.06
Model 258	70.59	70.02	69.73	69.58	66.00
Model 259	70.62	69.86	69.59	69.66	66.05
Model 260	70.45	69.84	69.49	69.64	66.24
Model 261	70.58	69.73	69.50	69.71	65.98
Model 262	70.76	69.84	69.70	69.65	66.03
Model 263	70.61	69.70	69.48	69.69	66.34
Model 264	70.30	69.73	69.73	69.53	66.07
Model 265	69.37	69.58	69.49	68.86	66.54
Model 266	69.73	69.55	69.55	68.96	66.40
Model 267	68.79	68.89	69.30	68.46	65.96
Model 268	68.93	69.92	68.97	68.77	66.50
Model 269	68.52	68.38	68.62	68.49	64.94
Model 270	65.89	65.90	65.91	65.92	62.09
Model 271	54.01	53.89	53.76	53.92	51.81
Model 272	53.35	52.65	53.09	53.34	51.80

Table 5.50. Precision values for Relief-F-based COVID-19 prediction models using different machine learning methods

Models	Precision (%)				
	MLP	TB	RBF	SVM	KMC
Model 256	65.33	66.23	66.43	66.76	64.94
Model 257	65.30	66.24	66.42	66.76	64.73
Model 258	65.88	66.40	66.31	66.86	64.80
Model 259	65.58	66.18	66.12	66.81	65.09
Model 260	64.85	66.34	65.99	66.78	65.27
Model 261	65.80	66.02	66.05	66.76	64.87
Model 262	65.37	66.18	66.29	66.59	65.00
Model 263	66.02	66.00	66.07	64.24	64.66
Model 264	64.97	66.03	66.23	66.19	64.92
Model 265	65.68	65.60	66.10	65.63	65.88
Model 266	65.58	65.62	66.18	65.70	65.58
Model 267	65.73	65.95	66.93	65.44	65.45
Model 268	65.69	66.94	66.39	65.97	65.97
Model 269	65.24	64.92	65.00	64.80	64.90
Model 270	70.92	70.65	70.74	70.39	63.20
Model 271	53.72	53.33	53.06	53.47	51.80
Model 272	52.89	52.48	52.95	53.23	51.79

Table 5.51. Recall values for Relief-F-based COVID-19 prediction models using different machine learning methods

Models	Recall (%)				
	MLP	TB	RBF	SVM	KMC
Model 256	87.96	81.18	80.36	77.19	70.44
Model 257	87.99	80.69	80.40	77.23	70.58
Model 258	85.46	81.04	80.22	77.64	70.04
Model 259	86.77	81.23	80.38	78.15	69.24
Model 260	89.32	80.55	80.43	78.14	69.42
Model 261	85.70	81.30	80.26	78.50	69.74
Model 262	88.29	81.13	80.17	78.87	69.45
Model 263	84.94	81.30	80.10	88.80	72.08
Model 264	88.06	81.27	80.51	79.83	69.92
Model 265	81.15	82.34	80.00	79.22	68.61
Model 266	83.04	82.09	79.94	79.33	69.01
Model 267	78.50	78.09	76.27	78.25	67.61
Model 268	79.26	78.70	76.84	77.52	68.13
Model 269	79.27	79.98	80.65	80.96	65.04
Model 270	53.87	54.39	54.24	54.96	57.87
Model 271	57.86	62.26	65.13	60.43	52.15
Model 272	61.30	56.17	55.43	54.99	52.07

Table 5.52. F1-score values for Relief-F-based COVID-19 prediction models using different machine learning methods

Models	F1-score (%)				
	MLP	TB	RBF	SVM	KMC
Model 256	74.97	72.95	72.73	71.59	67.58
Model 257	74.97	72.75	72.74	71.61	67.53
Model 258	74.40	72.99	72.60	71.85	67.32
Model 259	74.70	72.93	72.55	72.04	67.10
Model 260	75.14	72.76	72.50	72.02	67.28
Model 261	74.44	72.87	72.47	72.16	67.22
Model 262	75.12	72.90	72.58	72.21	67.15
Model 263	74.29	72.85	72.41	74.55	68.17
Model 264	74.78	72.86	72.68	72.37	67.33
Model 265	72.60	73.02	72.39	71.79	67.22
Model 266	73.28	72.94	72.41	71.88	67.25
Model 267	71.55	71.51	71.30	71.27	66.51
Model 268	71.84	72.34	71.23	71.28	67.03
Model 269	71.57	71.67	71.99	71.99	64.97
Model 270	61.23	61.46	61.40	61.73	60.42
Model 271	55.71	57.45	58.48	56.74	51.98
Model 272	56.79	54.26	54.16	54.10	51.93

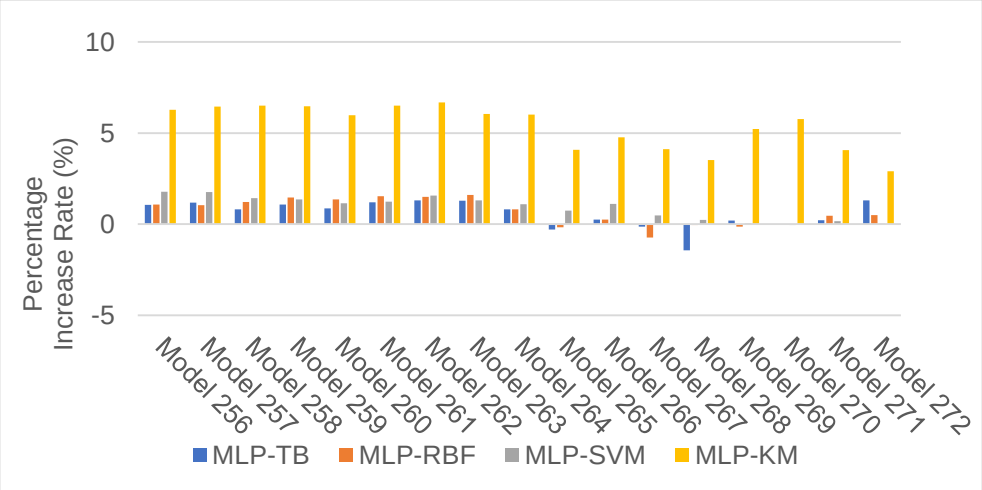


Figure 5.33. Percentage increase rates in accuracies of COVID-19 prediction for MLP-based models compared to accuracies obtained by the rest of the machine learning methods for Relief-F

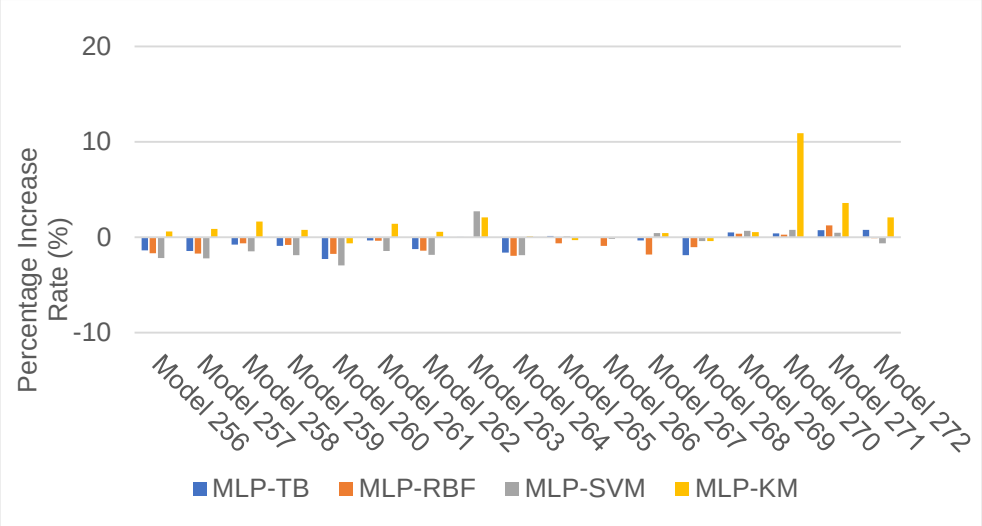


Figure 5.34. Percentage increase rates in precisions of COVID-19 prediction for MLP-based models compared to precisions obtained by the rest of the machine learning methods for Relief-F

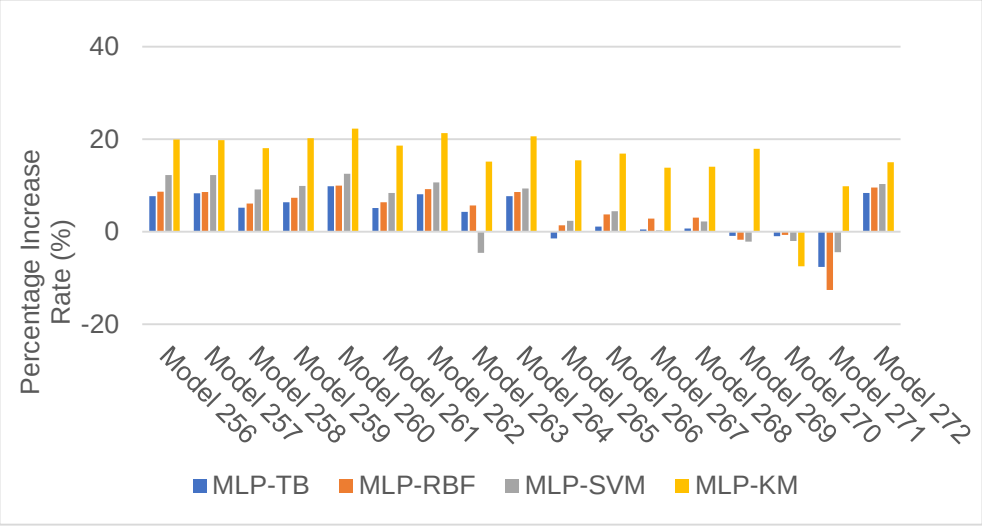


Figure 5.35. Percentage increase rates in recalls of COVID-19 prediction for MLP-based models compared to recalls obtained by the rest of the machine learning methods for Relief-F

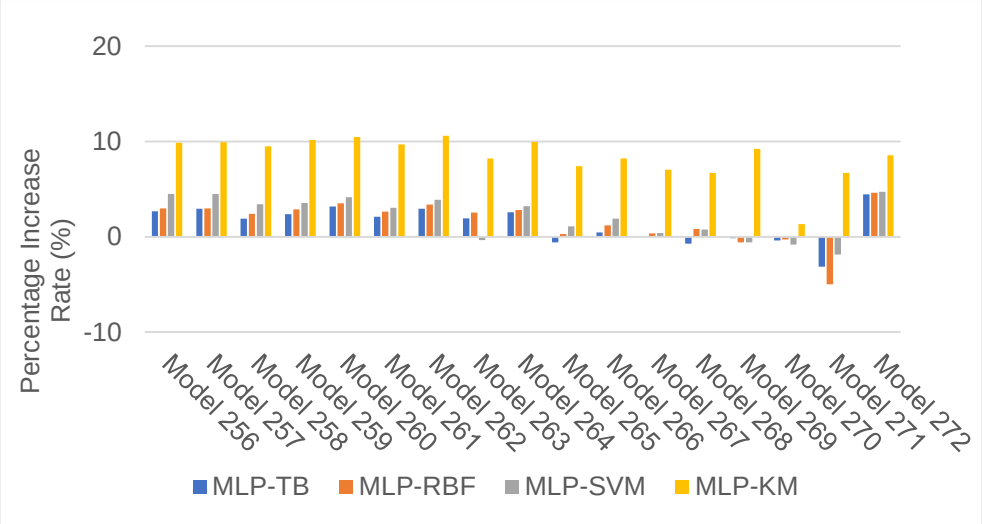


Figure 5.36. Percentage increase rates in F1-scores of COVID-19 prediction for MLP-based models compared to F1-scores obtained by the rest of the machine learning methods for Relief-F

5.17.2. mRMR Results

Table 5.50 through Table 5.53 show the accuracy, precision, recall, and F1-score values for COVID-19 prediction models including the mRMR combinations of the predictor variables, respectively. Figure 5.37 through 5.40 illustrates the percentage increase rates in accuracy, precision, recall, and F1-score values for COVID-19 prediction models.

Table 5.53. Accuracy values for mRMR-based COVID-19 prediction models using different machine learning methods

Models	Accuracy (%)				
	MLP	TB	RBF	SVM	KMC
Model 273	70.64	69.89	69.88	69.38	66.20
Model 274	70.56	70.04	69.72	69.38	65.81
Model 275	70.59	69.95	69.94	69.45	65.85
Model 276	70.73	70.03	69.84	69.76	66.32
Model 277	70.61	69.80	69.88	69.72	66.50
Model 278	70.63	69.88	69.85	70.47	66.86
Model 279	70.74	69.97	69.66	70.53	66.97
Model 280	70.67	69.95	69.67	70.56	66.54
Model 281	70.75	69.92	69.81	70.62	66.57
Model 282	69.56	69.38	69.22	69.12	66.88
Model 283	69.56	69.35	69.27	69.13	65.75
Model 284	69.14	68.99	69.03	68.89	65.98
Model 285	69.13	68.86	68.98	68.95	66.30
Model 286	69.03	68.88	69.12	68.99	65.75
Model 287	67.95	67.95	67.95	67.95	53.50
Model 288	67.17	67.17	67.17	67.17	50.00
Model 289	65.05	65.05	65.05	65.05	50.00

Table 5.54. Precision values for mRMR-based COVID-19 prediction models using different machine learning methods

Models	Precision (%)				
	MLP	**TB**	**RBF**	**SVM**	**KMC**
Model 273	65.33	66.23	66.43	66.76	64.94
Model 274	64.74	66.27	66.25	66.56	64.86
Model 275	65.66	66.29	66.43	66.44	65.33
Model 276	65.74	66.37	66.40	66.49	65.38
Model 277	65.70	66.28	66.35	66.46	65.05
Model 278	65.30	66.03	66.24	66.43	65.65
Model 279	65.90	66.23	66.10	66.42	65.68
Model 280	65.72	66.22	66.12	66.37	65.18
Model 281	66.00	66.11	66.29	66.40	65.19
Model 282	66.22	65.60	66.11	65.86	65.84
Model 283	65.85	65.57	66.23	65.38	65.30
Model 284	66.19	66.01	66.51	66.04	65.20
Model 285	66.05	66.03	66.35	66.12	65.84
Model 286	66.16	65.98	66.90	66.42	64.94
Model 287	65.77	65.77	65.77	65.77	58.49
Model 288	62.60	62.60	62.60	62.60	50.00
Model 289	68.42	68.42	68.42	68.42	50.00

Table 5.55. Recall values for mRMR-based COVID-19 prediction models using different machine learning methods

Models	Recall (%)				
	MLP	TB	RBF	SVM	KMC
Model 273	87.96	81.18	80.36	77.19	70.44
Model 274	90.29	81.61	80.38	77.88	68.97
Model 275	86.36	81.21	80.62	78.60	67.55
Model 276	86.57	81.24	80.31	79.67	69.37
Model 277	86.26	80.58	80.68	79.62	71.32
Model 278	88.04	81.87	80.97	82.77	70.73
Model 279	85.97	81.49	80.72	83.03	71.08
Model 280	86.40	81.43	80.71	83.34	71.01
Model 281	85.61	81.75	80.62	83.46	71.10
Model 282	79.87	81.50	78.85	79.39	70.17
Model 283	81.26	81.50	78.63	81.31	67.23
Model 284	78.27	78.29	76.65	77.78	68.58
Model 285	78.72	77.67	77.03	77.72	67.74
Model 286	77.91	77.96	75.67	76.83	68.48
Model 287	74.87	74.87	74.87	74.87	24.11
Model 288	85.29	85.29	85.29	85.29	68.45
Model 289	55.92	55.92	55.92	55.92	68.28

Table 5.56. F1-score values for mRMR-based COVID-19 prediction models using different machine learning methods

Models	F1-score (%)				
	MLP	TB	RBF	SVM	KMC
Model 273	74.97	72.95	72.73	71.59	67.58
Model 274	75.41	73.15	72.63	71.78	66.85
Model 275	74.60	72.99	72.84	72.01	66.42
Model 276	74.73	73.05	72.70	72.49	67.32
Model 277	74.59	72.74	72.82	72.44	68.04
Model 278	74.98	73.10	72.87	73.70	68.10
Model 279	74.61	73.07	72.68	73.80	68.27
Model 280	74.65	73.04	72.69	73.90	67.97
Model 281	74.53	73.10	72.76	73.96	68.02
Model 282	72.41	72.69	71.92	71.99	67.93
Model 283	72.75	72.67	71.90	72.48	66.25
Model 284	71.73	71.63	71.22	71.43	66.85
Model 285	71.83	71.38	71.29	71.45	66.78
Model 286	71.56	71.47	71.01	71.24	66.66
Model 287	70.02	70.02	70.02	70.02	34.15
Model 288	72.21	72.21	72.21	72.21	66.67
Model 289	61.54	61.54	61.54	61.54	66.67

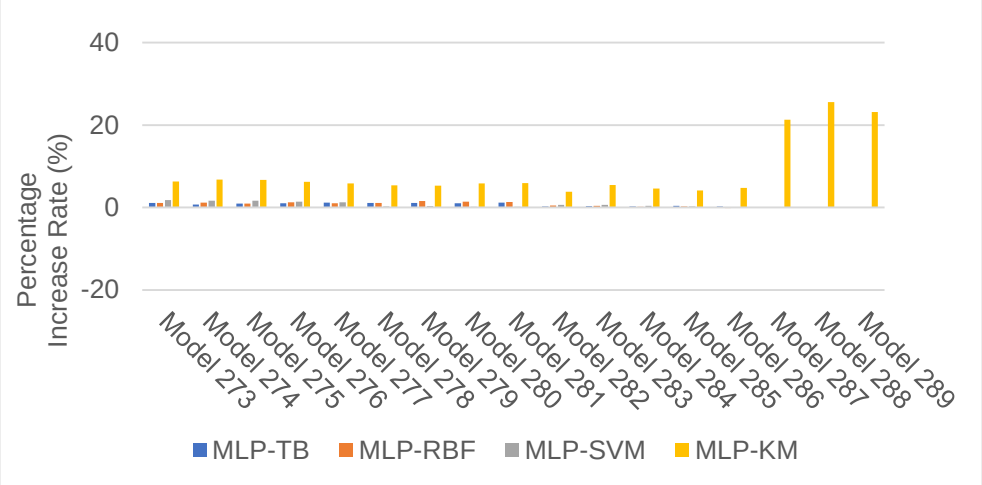


Figure 5.37. Percentage increase rates in accuracies of COVID-19 prediction for MLP-based models compared to accuracies obtained by the rest of the machine learning methods for mRMR

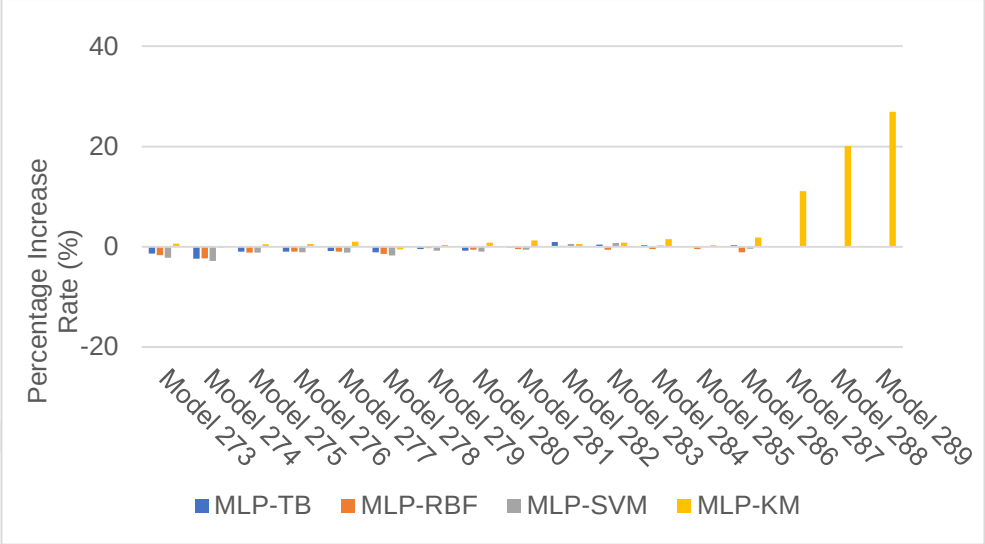


Figure 5.38. Percentage increase rates in precisions of COVID-19 prediction for MLP-based models compared to precisions obtained by the rest of the machine learning methods for mRMR

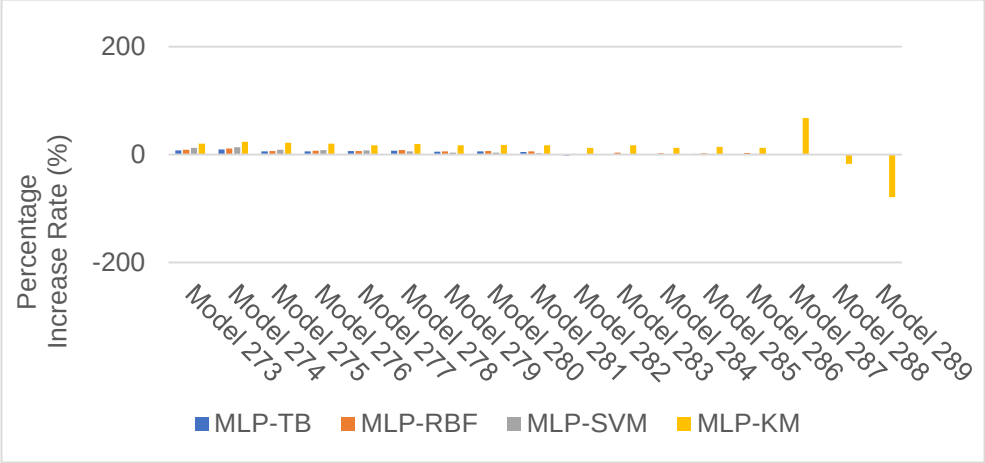


Figure 5.39. Percentage increase rates in recalls of COVID-19 prediction for MLP-based models compared to recalls obtained by the rest of the machine learning methods for mRMR

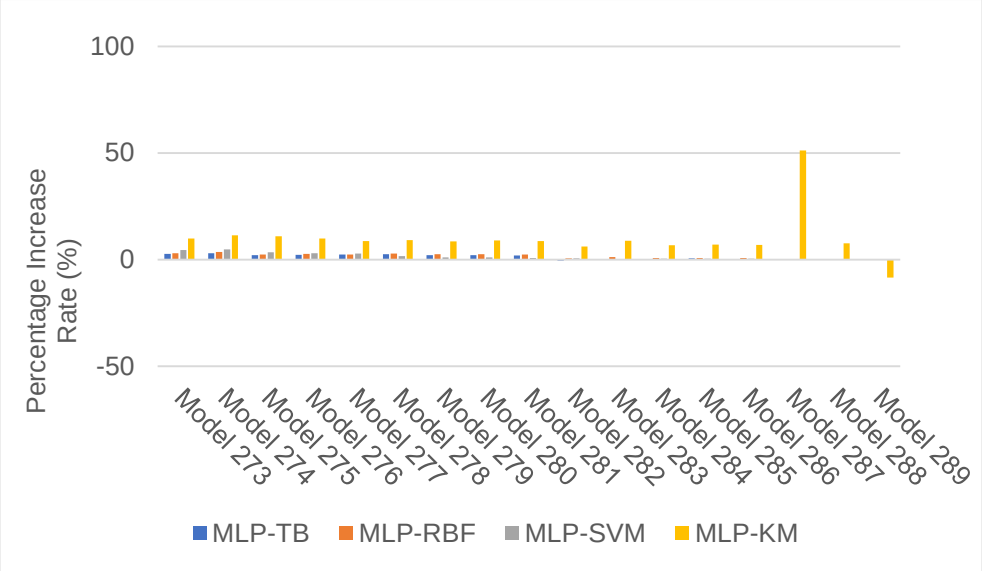


Figure 5.40. Percentage increase rates in F1-scores of COVID-19 prediction for MLP-based models compared to F1-scores obtained by the rest of the machine learning methods for mRMR

5.18. Discussion for Feature Selection-based Prediction Models

The following comments regarding the results can be made for the prediction of COVID-19 infection developed using Relief-F and mRMR scores of the predictor variables:

- According to mRMR and Relief-F, fever and age variables have been reported to be the most important variables for predicting COVID-19 infection, respectively. In contrast, the least essential variables for both feature selectors are the non-symptom variables, especially the comorbidities.
- The MLP-based Model 281 created using the mRMR feature selector, and the MLP-based Model 262 created using the Relief-F feature selector performed best in accuracy values with 70.75% and 70.76%, respectively.

- The MLP-based Model 274 created using the mRMR feature selector, and the MLP-based Model 260 created using the Relief-F feature selector achieved the highest recall values with 90.29% and 89.32%, respectively.
- The MLP-based Model 274 created using the mRMR feature selector, and the MLP-based Model 260 created using the Relief-F feature selector achieved the highest F1-scores with 75.41% and 75.14%, respectively.
- When the accuracy and F1-score values of the models created by using the Relief-F feature selector are compared, it is seen that the MLP-based models, in general, exhibit the best performance among the ML classifiers. The performance of TB-based, RBF-based, and SVM-based models are similar/comparable. In other words, there is no superiority of one classifier over the other two classifiers, but their accuracy and F1-score values are lower than the ones of MLP-based models. Finally, the KMC-based models produce the lowest accuracy and F1-score values. In contrast, when the recall values of the models created using the Relief feature selector are compared, the general order of ML classifiers leading from the best to the worst results is MLP, TB, RBF, SVM, and KMC.
- Similarly, when the accuracy and F1-score values of the models created by using the mRMR feature selector are compared, it is seen that again the MLP-based models, in general, produce the highest accuracy and F1-scores among the ML classifiers. The performance of TB-based, RBF-based, and SVM-based models are similar/comparable, occupying the second rank. Finally, the KMC-based models produce the lowest accuracy and F1-score values. In contrast, when the models' recall values created using the mRMR feature selector are compared, the

general order of ML classifiers leading from the best to the worst results is MLP, TB, RBF, SVM, and KMC. However, it should be noted that the SVM-based Model 278 through Model 284 exceptionally performed better than TB-based and RBF-based models.

- Models created by using the Relief-F feature selector produced average accuracies of 67.87%, 67.48%, 67.39%, 67.23%, 64.18%; average recall values of 79.93%, 76.69%, 75.96%, 75.29%, 66.55%; and average F1-scores of 71.02%, 70.03%, 69.80%, 69.48%, 64.94% for MLP, TB, RBF, SVM, KMC, respectively.
- Models created by using the mRMR feature selector produced average accuracies of 69.56%, 69.12%, 69.06%, 69.12%, 63.63%; average recall values of 79.88%, 77.67%, 76.21%, 78.41%, 68.55%; and average F1-scores of 70.97%, 70.96%, 70.05%, 71.97%, 66.73% for MLP, TB, RBF, SVM, and KMC, respectively. According to these results, it is seen that mRMR achieves slightly better results than Relief-F on average in terms of accuracy, recall, and F1-score values.
- Average training times of the MLP-based, RBF-based, SVM-based, TB-based, and KMC-based prediction models are appr. 16 min, 32 min, 13 h, 2 min, and 4 min, respectively. TB has been observed to be the most time-efficient classifier. It gives the second-best performance in the shortest time.

5.19. General Discussion

- When the performance of all models in the study is investigated, the following comment can be made: Though these models did not have significant performance differences, the general order of machine learning classifiers leading from best to worst results is MLP, TB, RBF,

SVM, and KMC. MLP-based models performed best in most cases, while KMC performed worst in most cases.

- For all methods used, at least appr. a 10% performance drop is observed in accuracy and precision metrics when the fever variable is removed from the model.
- It has been observed that feature selection-based models do not have any significant performance improvement over the other regular models.
- It has been observed that non-symptom variables (i.e., age, sex, race, obesity comorbidity) do not make any relative improvement or deterioration in the performance of the models, i.e., they have a negligible effect on predicting the COVID-19 infection.
- It has been observed that the mRMR feature selection method performs slightly better than the Relief-F feature selection method. Also, it has been observed that mRMR is better at finding the relevance order of the predictor variables than the Relief-F feature selection method. Finally, the mRMR feature selector puts more importance on symptomatic variables such as fever, cough, and headache. On the other hand, Relief-F puts more importance on non-symptomatic variables such as age, race, or cardio comorbidity.
- Among all MLP-based models, Model 34 including the common variables with diabetes and smoking comorbidities exhibits the best accuracy value with 70.98%.
- Among all MLP-based models, Model 270 including age, race and fever exhibits the best precision value with 70.92%.
- Among all MLP-based models, Model 63 including the common variables with race, cardio, and renal comorbidities exhibits the best recall value with 91.31%.

- Among all MLP-based models, Model 63 including the common variables with race, cardio, and renal comorbidities exhibits the best F1-score value with 75.61%.
- Among all TB-based models, Model 151 including the common variables with lung, cardio, diabetes, and smoking comorbidities exhibits the best accuracy value of 70.27%.
- Among all TB-based models, Model 270 including age, race, and fever variables exhibits the best precision value with 70.65%.
- Among all TB-based models, Model 288 including fever and cough variables exhibits the best recall value with 85.29%.
- Among all TB-based models, Model 154 including the common variables with lung, renal, diabetes, and smoking comorbidities exhibits the best F1-score value with 73.47%.
- Among all RBF-based models, Model 68 including the common variables with race, renal, and smoking comorbidities exhibits the best accuracy value with 70.04% .
- Among all RBF-based models, Model 270 including age, race, and fever variables exhibits the best precision value among all RBF-based models with 70.74%.
- Among all RBF-based models, Model 288 including fever and cough variables exhibits the best recall value with 85.29%.
- Among all RBF-based models, Model 155 including the common variables with lung, renal, diabetes, and obesity comorbidities exhibits the best F1-score value with 73.02%.
- Among all SVM-based models, Model 281 including fever, cough, headache, age, breathing difficulty, obesity comorbidity, throat pain, diarrhea, and runny nose variables exhibits the best accuracy value with 70.62%.

- Among all SVM-based models, Model 270 including age, race, and fever variables exhibits the best precision value with 70.39%.
- Among all SVM-based models, Model 239 including the common variables with pregnant, cardio, renal, diabetes, smoking, and obesity comorbidities exhibits the best recall value with 89.23%.
- Among all SVM-based models, Model 35 including the common variables diabetes and obesity comorbidities exhibits the best F1-score value with 74.80%.
- Among all KMC-based models, Model 226 including the common variables with pregnant, race, lung, renal, diabetes, and obesity comorbidities exhibits the best accuracy value with 67.71%.
- Among all KMC-based models, Model 25 including the common variables with lung and smoking comorbidities exhibits the best precision value with 65.99%.
- Among all KMC-based models, Model 226 including the common variables with pregnant, race, lung, renal, diabetes, and obesity comorbidities exhibits the best recall value with 74.59%.
- Among all KMC-based models, Model 226 including the common variables with pregnant, race, lung, renal, diabetes, and obesity comorbidities exhibits the best F1-score value with 69.79%.

5.20. Comparing Results with Other Studies in the Literature

An interesting question is about how the accuracy, precision, recall, and F1-score of the best performing MLP-based models for each respective metric (i.e., Model 34, Model 63, and Model 270) compares to the metrics of regular models in the literature.

The best-performing models for each metric have been compared with the metrics of models proposed in the literature. The best performing models for

accuracy, precision, recall, and F1-score are Model 34, Model 270, Model 63, and Model 63, respectively. It should also be considered that for the most of studies listed in Table 2.1 through Table 2.8 in Section 2., a direct comparison with the prediction results presented in this study is impossible because (a) the datasets used by each study contains a different size of data; (b) the datasets used by each study have different data features that most frequently belong to specialized target audiences, such as different age group patients, patients with different nationalities, patients from different locations; (c) each study uses various different prediction models and predictors, and (d) each study utilizes various validation techniques.

However, apart from such incomparable studies and prediction results, there are two significant studies in the literature (Fayyoumi et al., 2020; Santana et al., 2021) that have the same purpose as this study (i.e., COVID-19 prediction using symptoms and other characteristics of the patients). This situation allows us to make an indirect comparison between the metrics of the prediction models in this study and metrics from the study in (Fayyoumi et al., 2020) and (Santana et al., 2021). Particularly, the studies in (Fayyoumi et al., 2020) and (Santana et al., 2021) found MLP to be the best-performing ML classifier compared to other ML classifiers as this study. The highest performing models' accuracy, precision, recall, and F1-scores of these two studies are illustrated in Table 5.54.

There are two main reasons why this study yielded lower performance metrics compared to these two studies in the literature. First, during the preprocessing phase, asymptomatic cases have only partially been eliminated from the dataset, but other studies eliminated all asymptomatic patients. Second, this study uses a more extensive dataset than other studies. Particularly, (Fayyoumi et al., 2020) and (Santana et al., 2021) only used 105 and 2674 cases, respectively, whereas in this study, 20.000 cases were utilized.

Table 5.57. Comparison of the results with previous works

Study	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
(Fayyoumi et al., 2020)	MLP	91.62	90.00	87.80	88.89
(Santana et al., 2021)	MLP	97.86	97.68	98.05	97.86
This Study - Model 34	MLP	70.98	66.04	86.37	74.85
This Study - Model 63	MLP	70.55	64.52	91.31	75.61
This Study – Model 270	MLP	65.91	70.99	53.81	61.22

6. CONCLUSION AND FUTURE WORKS

COVID-19 is a highly infectious respiratory virus that infected more than 465 million people and was declared a pandemic by WHO in March 2020. COVID-19 patients may have asymptomatic, mild, or severe symptoms and are most commonly verified through PCR tests. Many studies about other diseases have already been conducted, but the different characteristics and spread rate of COVID-19 make it unique from other diseases such as h1n1, swine flu, or influenza. Many governments and businesses take measures almost never taken before, such as mandatory mask-wearing, curfews, and temporary closure of non-emergency institutions. To limit the spread of COVID-19 and plan ahead, developing new prediction models becomes necessary so governments and businesses may avoid problems such as inadequate supplies, overfilled facilities, and insufficient personnel.

This study predicted the COVID-19 infection by developing two types of prediction models, namely regular and feature selection-based prediction models. MLP, TB, RBF, SVM, and KMC, individually combined with mRMR and Relief-F, have been used to construct new feature selection-based COVID-19 prediction models and discern the influential variables for the prediction of COVID-19 infection. Also, further models were created using the unique to octuple combinations of the predictor variables. Finally, the accuracy, precision, recall, and F1-score metrics have been used to assess the models' performance, whereas the generalization errors of the models were evaluated using 10-fold cross-validation.

The utilized dataset contains information related to 20.000 patients (i.e., 10.000 positives, 10.000 negatives) and includes several personal, symptomatic, and non-symptomatic variables. Performing 10-fold cross-validation on the dataset, the performance of the models has been assessed by calculating the accuracy, recall, and F1-scores. The results show that the average performance of mRMR is slightly better

than Relief-F. Furthermore, mRMR is more successful than the Relief-F algorithm in finding the relative relevance order of the predictor variables to predict COVID-19 infection. Finally, the mRMR feature selector emphasizes symptomatic variables such as fever and cough, whereas the Relief-F algorithm emphasizes non-symptomatic variables such as age and race. In addition, when the performance of the ML classifiers is examined, it is observed that MLP outperforms every ML classifier in accuracy, recall, and F1-score metrics. MLP has proven itself as a relatively fast and reliable ML method for this prediction field. On the other hand, KMC exhibited the worst performance in every metric. In addition, TB has the shortest training time among all ML classifiers and has the second best performance in the study.

In the future, this study could be extended in multiple directions. More studies on more datasets with different characteristics are needed to generalize the advantage of one ML classifier over the others in this field. A comprehensive study on datasets based on Turkish patients is also missing in this field. Furthermore, the existence of vaccines creates a need for further investigations in this field. For example, the vaccines may affect the importance of the variables because patients may have milder symptoms compared to non-vaccinated patients, thus making some of the predictor variables ineffective.

REFERENCES

Acid, S., De Campos, L. M., & Fernández, M. 2011. Minimum redundancy maximum relevancy versus score-based methods for learning Markov boundaries. In 2011 11th International Conference on Intelligent Systems Design and Applications IEEE, 619-623.

Aditya Satrio, C. B., Darmawan, W., Nadia, B. U., & Hanafiah, N. 2021. Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET. In Procedia Computer Science, 179:524–532.

Ahmed, S. Z., 2020. Analysis and forecasting the outbreak of covid-19 in Ethiopia using Machine learning. European Journal of Computer Science and Information Technology, 8(4):1-13.

Althnian, A., Elwafa, A. A., Aloboud, N., Alrasheed, H., & Kurdi, H., 2020. Prediction of COVID-19 Individual Susceptibility using Demographic Data: A Case Study on Saudi Arabia. In Procedia Computer Science, 177:379–386.

Ardabili, S. F., Mosavi, A., Ghamisi, P., Ferdinand, F., Varkonyi-Koczy, A. R., Reuter, U., Rabczuk, T., & Atkinson, P. M., 2020. COVID-19 Outbreak Prediction with Machine Learning. Algorithms, 13(10):249.

Ayoobi, N., Sharifrazi, D., Alizadehsani, R., Shoeibi, A., Gorriz, J. M., Moosaei, H., Khosravi, A., Nahavandi, S., Gholamzadeh Chofreh, A., Goni, F. A., Klemeš, J. J., & Mosavi, A., 2021. Time series forecasting of new cases and new deaths rate for COVID-19 using deep learning methods. In Results in Physics, 27:104495.

Bala, S., 2021. COVID-19 Outbreak Prediction Analysis using Machine Learning. In International Journal for Research in Applied Science and Engineering Technology, 9(1):1–7.

Bhavsar, H., & Panchal, M. H., 2012. A review on support vector machine for data classification. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 1(10):185-189.

Ceylan, Z., 2020. Estimation of COVID-19 prevalence in Italy, Spain, and France. In Science of The Total Environment, 729:138817.

Ciotti, M., Ciccozzi, M., Terrinoni, A., Jiang, W.-C., Wang, C.-B., & Bernardini, S., 2020. The COVID-19 pandemic. In Critical Reviews in Clinical Laboratory Sciences, 57(6):365–388.

Da Silva, R. G., Ribeiro, M. H. D. M., Mariani, V. C., & Coelho, L. dos S., 2020. Forecasting Brazilian and American COVID-19 cases based on artificial intelligence coupled with climatic exogenous variables. In Chaos, Solitons & Fractals, 139:110027.

Dash, Ch. S. K., Behera, A. K., Dehuri, S., & Cho, S.-B., 2016. Radial basis function neural networks: a topical state-of-the-art survey. In Open Computer Science, 6(1):33–63.

De Souza, F. S. H., Hojo-Souza, N. S., Dos Santos, E. B., Da Silva, C. M., & Guidoni, D. L., 2021. Predicting the Disease Outcome in COVID-19 Positive Patients Through Machine Learning: A Retrospective Cohort Study with Brazilian Data. In Frontiers in Artificial Intelligence, 4.

Djeddou, M., Hameed, I. A., Nejatian, A., & Loukam, I., 2020. Predictive Modelling of COVID-19 New Cases in Algeria using An Extreme Learning Machines (ELM). Cold Spring Harbor Laboratory. Preprint doi: 10.1101/2020.09.28.20203299

Fang, L., Wang, D., & Pan, G. (2020). Analysis and Estimation of COVID-19 Spreading in Russia Based on ARIMA Model. In SN Comprehensive Clinical Medicine, 2(12):2521–2527.

Fayyoumi, E., Idwan, S., & AboShindi, H., 2020. Machine Learning and Statistical Modelling for Prediction of Novel COVID-19 Patients Case Study: Jordan. In International Journal of Advanced Computer Science and Applications, 11(5).

Friedman, J. H. 2002. Stochastic gradient boosting. In Computational Statistics & Data Analysis, 38(4):367–378.

Gebretensae, Y. A., & Asmelash, D., 2021. Trend Analysis and Forecasting the Spread of COVID-19 Pandemic in Ethiopia Using Box–Jenkins Modeling Procedure. In International Journal of General Medicine, 14:1485–1498.

Gupta, A. K., Singh, V., Mathur, P., & Travieso-Gonzalez, C. M., 2020. Prediction of COVID-19 pandemic measuring criteria using support vector machine, prophet and linear regression models in Indian scenario. In Journal of Interdisciplinary Mathematics, 24(1):89–108.

Hassan, A. H. M., Qasem, A. A. M., Abdalla, W. F. M., & Elhassan, O. H., 2021. Visualization & Prediction of COVID-19 Future Outbreak by Using Machine Learning. In International Journal of Information Technology and Computer Science, 13(3):16–32.

Jojoa, M., & Garcia-Zapirain, B., 2020. Forecasting COVID 19 Confirmed Cases Using Machine Learning: The Case of America. MDPI AG. Preprint doi: 10.20944/preprints202009.0228.v1

Luo, J., Zhang, Z., Fu, Y., & Rao, F., 2021. Time series prediction of COVID-19 transmission in America using LSTM and XGBoost algorithms. In Results in Physics, 27:104462.

Maloney, K. O., Schmid, M., & Weller, D. E., 2011. Applying additive modelling and gradient boosting to assess the effects of watershed and reach characteristics on riverine assemblages. In Methods in Ecology and Evolution, 3(1):116–128.

Marzouk, M., Elshaboury, N., Abdel-Latif, A., & Azab, S., 2021. Deep learning model for forecasting COVID-19 outbreak in Egypt. In Process Safety and Environmental Protection, 153:363–375.

Misra, D., 2019. Mish: A Self Regularized Non-Monotonic Activation Function. arxiv. Preprint doi: 10.48550/arXiv.1908.08681

Moreau, V. H., 2020. Forecast predictions for the COVID-19 pandemic in Brazil by statistical modeling using the Weibull distribution for daily new cases and deaths. In Brazilian Journal of Microbiology, 51(3):1109–1115.

Onder, H., 2020. Short-term Forecasts of the COVID-19 Epidemic in Turkey: March 16–28, 2020. Black Sea Journal of Health Science, 3(2):27-30.

Open Data from the State of Espirito Santo. Data on COVID-19 pandemic, URL: https://dados.es.gov.br/-dataset/dados-sobre-pandemia-covid-19/resource-/38cc5066-020d-4c5a-b4c0-e9f690deb6d4. Last Access: 24.05.2021

Pinter, G., Felde, I., Mosavi, A., Ghamisi, P., & Gloaguen, R., 2020. COVID-19 Pandemic Prediction for Hungary: A Hybrid Machine Learning Approach. In Mathematics, 8(6):890.

Pourghasemi, H. R., Pouyan, S., Farajzadeh, Z., Sadhasivam, N., Heidari, B., Babaei, S., & Tiefenbacher, J. P., 2020. Assessment of the outbreak risk, mapping and infection behavior of COVID-19: Application of the autoregressive integrated-moving average (ARIMA) and polynomial models. In J. Zhang (Ed.), PLOS ONE, 15(7):e0236238.

Prakash, K. B., 2020. Analysis, Prediction and Evaluation of COVID-19 Datasets using Machine Learning Algorithms. In International Journal of Emerging Trends in Engineering Research, 8(5):2199–2204.

Punn, N. S., Sonbhadra, S. K., & Agarwal, S., 2020. COVID-19 Epidemic Analysis using Machine Learning and Deep Learning Algorithms. Cold Spring Harbor Laboratory. Preprint doi: 10.1101/2020.04.08.20057679

Ramchoun, H., Amine, M., Idrissi, J., Ghanou, Y., & Ettaouil, M., 2016. Multilayer Perceptron: Architecture Optimization and Training. In International Journal of Interactive Multimedia and Artificial Intelligence, 4(1):26.

Robnik-Šikonja, M., & Kononenko, I., 2003. Theoretical and empirical analysis of ReliefF and RReliefF. Machine learning, 53(1):23-69.

Rustam, F., Reshi, A. A., Mehmood, A., Ullah, S., On, B.-W., Aslam, W., & Choi, G. S., 2020. COVID-19 Future Forecasting Using Supervised Machine Learning Models. In IEEE Access, 8:101489–101499.

Saba, A. I., & Elsheikh, A. H., 2020. Forecasting the prevalence of COVID-19 outbreak in Egypt using nonlinear autoregressive artificial neural networks. In Process Safety and Environmental Protection 141:1–8.

Sahin, M., 2020. Forecasting COVID-19 cases based on mobility. In MANAS Journal of Engineering 8(2):144–150.

Sherrod, P., 2016. DTREG Predictive Modelling Software. Networks, 12:28–35.

Shukla, S., & Naganna, S., 2014. A review on K-means data clustering approach. International Journal of Information and Computation Technology, 4(17):1847-1860.

Takele, R., 2020. Stochastic modelling for predicting COVID-19 prevalence in East Africa Countries. In Infectious Disease Modelling, 5:598–607.

Tamhane, R., & Mulge, S., 2020. Prediction of COVID-19 outbreak using machine learning. International Research Journal of Engineering and Technology (IRJET), 7(5).

Tuli, S., Tuli, S., Tuli, R., & Gill, S. S., 2020. Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing. In Internet of Things, 11:100222.

Viana dos Santos Santana, Í., CM da Silveira, A., Sobrinho, Á., Chaves e Silva, L., Dias da Silva, L., Santos, D. F. S., Gurjão, E. C., & Perkusich, A., 2021. Classification Models for COVID-19 Test Prioritization in Brazil: Machine Learning Approach. In Journal of Medical Internet Research, 23(4):e27293.

Wollenstein-Betech, S., Cassandras, C. G., & Paschalidis, I. Ch., 2020. Personalized predictive models for symptomatic COVID-19 patients using basic preconditions: Hospitalizations, mortality, and the need for an ICU or ventilator. In International Journal of Medical Informatics, 142:104258.

Worldometers. COVID Live, URL: https://www.-worldometers.info/coronavirus/ Last Access: 15.05.2022.

Yu, C.-S., Chang, S.-S., Chang, T.-H., Wu, J. L., Lin, Y.-J., Chien, H.-F., & Chen, R.-J., 2021. A COVID-19 Pandemic Artificial Intelligence–Based System with

Deep Learning Forecasting and Automatic Statistical Data Acquisition: Development and Implementation Study. In Journal of Medical Internet Research 23(5):e27806.

CURRICULUM VITAE

Umut Ahmet ÇETIN. He completed his high school education at Ismail Safa Özler Anatolian High School, Adana in 2014. He received his Computer Engineering degree from Iskenderun Technical University, Hatay, in 2020.

APPENDIX

A.1. Prediction Models Created Using Unique Combinations

By using the unique combinations of the predictor variables, a total of eight prediction models have been formed, respectively. Table A.5.1 shows the overview of COVID-19 prediction models along with unique predictor variables.

Table A.5.1. Overview of COVID-19 prediction models along with unique predictor variables

Models	**Predictor Variables**
Model 1	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant
Model 2	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race
Model 3	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Lung Comorbidity
Model 4	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Cardio Comorbidity
Model 5	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Renal Comorbidity
Model 6	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Diabetes Comorbidity
Model 7	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Smoking Comorbidity
Model 8	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Obesity Comorbidity

A.2. Prediction Models Created Using Double Combinations

By using double combinations of the predictor variables, a total of twenty-eight prediction models have been formed, respectively. Table A.5.2 through Table A.5.4 show the overview of COVID-19 prediction models along with double combinations of the predictor variables.

Table A.5.2. Overview of COVID-19 prediction models along with double combinations of the predictor variables. – Part 1

Models	**Predictor Variables**
Model 9	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race
Model 10	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Lung Comorbidity
Model 11	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Cardio Comorbidity
Model 12	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Renal Comorbidity
Model 13	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Diabetes Comorbidity
Model 14	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Smoking Comorbidity
Model 15	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Obesity Comorbidity
Model 16	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Lung Comorbidity
Model 17	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Cardio Comorbidity
Model 18	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Renal Comorbidity
Model 19	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Diabetes Comorbidity

Table A.5.3. Overview of COVID-19 prediction models along with double combinations of the predictor variables – Part 2

Models	**Predictor Variables**
Model 20	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Smoking Comorbidity
Model 21	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Obesity Comorbidity
Model 22	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Lung Comorbidity, Cardio Comorbidity
Model 23	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Lung Comorbidity, Renal Comorbidity
Model 24	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Lung Comorbidity, Diabetes Comorbidity
Model 25	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Lung Comorbidity, Smoking Comorbidity
Model 26	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Lung Comorbidity, Obesity Comorbidity
Model 27	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Cardio Comorbidity, Renal Comorbidity
Model 28	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Cardio Comorbidity, Diabetes Comorbidity
Model 29	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Cardio Comorbidity, Smoking Comorbidity
Model 30	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Cardio Comorbidity, Obesity Comorbidity

Table A.5.4. Overview of COVID-19 prediction models along with double combinations of the predictor variables – Part 3

Models	**Predictor Variables**
Model 31	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Renal Comorbidity, Diabetes Comorbidity
Model 32	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Renal Comorbidity, Smoking Comorbidity
Model 33	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Renal Comorbidity, Obesity Comorbidity
Model 34	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Diabetes Comorbidity, Smoking Comorbidity
Model 35	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Diabetes Comorbidity, Obesity Comorbidity
Model 36	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Smoking Comorbidity, Obesity Comorbidity

A.3. Prediction Models Created Using Triple Combinations

By using triple combinations of the predictor variables, a total of fifty-six prediction models have been formed, respectively. Table A.5.5 through Table A.5.11 show the overview of COVID-19 prediction models along with triple combinations of the predictor variables.

Table A.5.5. Overview of COVID-19 prediction models along with triple combinations of the predictor variables. – Part 1

Models	**Predictor Variables**
Model 37	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Lung Comorbidity
Model 38	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Cardio Comorbidity
Model 39	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Renal Comorbidity
Model 40	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Diabetes Comorbidity
Model 41	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Smoking Comorbidity
Model 42	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Obesity Comorbidity
Model 43	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Lung Comorbidity, Cardio Comorbidity
Model 44	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Lung Comorbidity, Renal Comorbidity
Model 45	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Lung Comorbidity, Diabetes Comorbidity

Table A.5.6. Overview of COVID-19 prediction models along with triple combinations of the predictor variables. – Part 2

Models	**Predictor Variables**
Model 46	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Lung Comorbidity, Smoking Comorbidity
Model 47	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Lung Comorbidity, Obesity Comorbidity
Model 48	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Cardio Comorbidity, Renal Comorbidity
Model 49	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Cardio Comorbidity, Diabetes Comorbidity
Model 50	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Cardio Comorbidity, Smoking Comorbidity
Model 51	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Cardio Comorbidity, Obesity Comorbidity
Model 52	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Renal Comorbidity, Diabetes Comorbidity
Model 53	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Renal Comorbidity, Smoking Comorbidity
Model 54	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Renal Comorbidity, Obesity Comorbidity
Model 55	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Diabetes Comorbidity, Smoking Comorbidity

Table A.5.7. Overview of COVID-19 prediction models along with triple combinations of the predictor variables. – Part 3

Models	**Predictor Variables**
Model 56	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Diabetes Comorbidity, Obesity Comorbidity
Model 57	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Smoking Comorbidity, Obesity Comorbidity
Model 58	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Lung Comorbidity, Cardio Comorbidity
Model 59	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Lung Comorbidity, Renal Comorbidity
Model 60	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Lung Comorbidity, Diabetes Comorbidity
Model 61	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Lung Comorbidity, Smoking Comorbidity
Model 62	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Lung Comorbidity, Obesity Comorbidity
Model 63	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Cardio Comorbidity, Renal Comorbidity

Table A.5.8. Overview of COVID-19 prediction models along with triple combinations of the predictor variables. – Part 4

Models	Predictor Variables
Model 64	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Cardio Comorbidity, Diabetes Comorbidity
Model 65	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Cardio Comorbidity, Smoking Comorbidity
Model 66	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Cardio Comorbidity, Obesity Comorbidity
Model 67	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Renal Comorbidity, Diabetes Comorbidity
Model 68	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Renal Comorbidity, Smoking Comorbidity
Model 69	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Renal Comorbidity, Obesity Comorbidity
Model 70	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Diabetes Comorbidity, Smoking Comorbidity
Model 71	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Diabetes Comorbidity, Obesity Comorbidity

Table A.5.9. Overview of COVID-19 prediction models along with triple combinations of the predictor variables. – Part 5

Models	**Predictor Variables**
Model 72	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Smoking Comorbidity, Obesity Comorbidity
Model 73	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Lung Comorbidity, Cardio Comorbidity, Renal Comorbidity
Model 74	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Lung Comorbidity, Cardio Comorbidity, Diabetes Comorbidity
Model 75	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Lung Comorbidity, Cardio Comorbidity, Smoking Comorbidity
Model 76	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Lung Comorbidity, Cardio Comorbidity, Obesity Comorbidity

Table A.5.10. Overview of COVID-19 prediction models along with triple combinations of the predictor variables. – Part 6

Models	Predictor Variables
Model 77	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Lung Comorbidity, Renal Comorbidity, Diabetes Comorbidity
Model 78	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Lung Comorbidity, Renal Comorbidity, Smoking Comorbidity
Model 79	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Lung Comorbidity, Renal Comorbidity, Obesity Comorbidity
Model 80	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Lung Comorbidity, Diabetes Comorbidity, Smoking Comorbidity
Model 81	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Lung Comorbidity, Diabetes Comorbidity, Obesity Comorbidity
Model 82	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Lung Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 83	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Cardio Comorbidity, Renal Comorbidity, Diabetes Comorbidity
Model 84	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Cardio Comorbidity, Renal Comorbidity, Smoking Comorbidity
Model 85	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Cardio Comorbidity, Renal Comorbidity, Obesity Comorbidity
Model 86	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Cardio Comorbidity, Diabetes Comorbidity, Smoking Comorbidity

Table A.5.11. Overview of COVID-19 prediction models along with triple combinations of the predictor variables. – Part 7

Models	Predictor Variables
Model 87	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Cardio Comorbidity, Diabetes Comorbidity, Obesity Comorbidity
Model 88	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Cardio Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 89	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Renal Comorbidity, Diabetes Comorbidity, Smoking Comorbidity
Model 90	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Renal Comorbidity, Diabetes Comorbidity, Obesity Comorbidity
Model 91	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Renal Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 92	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Diabetes Comorbidity, Smoking Comorbidity, Obesity Comorbidity

A.4. Prediction Models Created Using Quadruple Combinations

By using quadruple combinations of the predictor variables, a total of seventy prediction models have been formed, respectively. Table A.5.12 through Table A.5.19 show the overview of COVID-19 prediction models along with quadruple combinations of the predictor variables.

Table A.5.12. Overview of COVID-19 prediction models along with quadruple combinations of the predictor variables – Part 1

Models	**Predictor Variables**
Model 93	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Lung Comorbidity, Cardio Comorbidity
Model 94	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Lung Comorbidity, Renal Comorbidity
Model 95	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Lung Comorbidity, Diabetes Comorbidity
Model 96	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Lung Comorbidity, Smoking Comorbidity
Model 97	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Lung Comorbidity, Obesity Comorbidity
Model 98	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Cardio Comorbidity, Renal Comorbidity
Model 99	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Cardio Comorbidity, Diabetes Comorbidity
Model 100	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Cardio Comorbidity, Smoking Comorbidity

Table A.5.13. Overview of COVID-19 prediction models along with quadruple combinations of the predictor variables – Part 2

Models	**Predictor Variables**
Model 101	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Cardio Comorbidity, Obesity Comorbidity
Model 102	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Renal Comorbidity, Diabetes Comorbidity
Model 103	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Renal Comorbidity, Smoking Comorbidity
Model 104	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Renal Comorbidity, Obesity Comorbidity
Model 105	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Diabetes Comorbidity, Smoking Comorbidity
Model 106	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Diabetes Comorbidity, Obesity Comorbidity
Model 107	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Smoking Comorbidity, Obesity Comorbidity
Model 108	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Lung Comorbidity, Cardio Comorbidity, Renal Comorbidity
Model 109	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Lung Comorbidity, Cardio Comorbidity, Diabetes Comorbidity
Model 110	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Lung Comorbidity, Cardio Comorbidity, Smoking Comorbidity

Table A.5.14. Overview of COVID-19 prediction models along with quadruple combinations of the predictor variables – Part 3

Models	**Predictor Variables**
Model 111	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Lung Comorbidity, Cardio Comorbidity, Obesity Comorbidity
Model 112	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Lung Comorbidity, Renal Comorbidity, Diabetes Comorbidity
Model 113	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Lung Comorbidity, Renal Comorbidity, Smoking Comorbidity
Model 114	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Lung Comorbidity, Renal Comorbidity, Obesity Comorbidity
Model 115	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Lung Comorbidity, Diabetes Comorbidity, Smoking Comorbidity
Model 116	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Lung Comorbidity, Diabetes Comorbidity, Obesity Comorbidity
Model 117	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Lung Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 118	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Cardio Comorbidity, Renal Comorbidity, Diabetes Comorbidity
Model 119	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Cardio Comorbidity, Renal Comorbidity, Smoking Comorbidity
Model 120	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Cardio Comorbidity, Renal Comorbidity, Obesity Comorbidity

Table A.5.15. Overview of COVID-19 prediction models along with quadruple combinations of the predictor variables – Part 4

Models	Predictor Variables
Model 121	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Cardio Comorbidity, Diabetes Comorbidity, Smoking Comorbidity
Model 122	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Cardio Comorbidity, Diabetes Comorbidity, Obesity Comorbidity
Model 123	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Cardio Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 124	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Renal Comorbidity, Diabetes Comorbidity, Smoking Comorbidity
Model 125	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Renal Comorbidity, Diabetes Comorbidity, Obesity Comorbidity
Model 126	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Renal Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 127	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Diabetes Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 128	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Lung Comorbidity, Cardio Comorbidity, Renal Comorbidity
Model 129	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Lung Comorbidity, Cardio Comorbidity, Diabetes Comorbidity
Model 130	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Lung Comorbidity, Cardio Comorbidity, Smoking Comorbidity

Table A.5.16. Overview of COVID-19 prediction models along with quadruple combinations of the predictor variables – Part 5

Models	Predictor Variables
Model 131	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Lung Comorbidity, Cardio Comorbidity, Obesity Comorbidity
Model 132	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Lung Comorbidity, Renal Comorbidity, Diabetes Comorbidity
Model 133	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Lung Comorbidity, Renal Comorbidity, Smoking Comorbidity
Model 134	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Lung Comorbidity, Renal Comorbidity, Obesity Comorbidity
Model 135	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Lung Comorbidity, Diabetes Comorbidity, Smoking Comorbidity
Model 136	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Lung Comorbidity, Diabetes Comorbidity, Obesity Comorbidity
Model 137	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Lung Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 138	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Cardio Comorbidity, Renal Comorbidity, Diabetes Comorbidity
Model 139	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Cardio Comorbidity, Renal Comorbidity, Smoking Comorbidity

Table A.5.17. Overview of COVID-19 prediction models along with quadruple combinations of the predictor variables – Part 6

Models	Predictor Variables
Model 140	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Cardio Comorbidity, Renal Comorbidity, Obesity Comorbidity
Model 141	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Cardio Comorbidity, Diabetes Comorbidity, Smoking Comorbidity
Model 142	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Cardio Comorbidity, Diabetes Comorbidity, Obesity Comorbidity
Model 143	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Cardio Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 144	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Renal Comorbidity, Diabetes Comorbidity, Smoking Comorbidity
Model 145	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Renal Comorbidity, Diabetes Comorbidity, Obesity Comorbidity
Model 146	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Renal Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 147	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Diabetes Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 148	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Lung Comorbidity, Cardio Comorbidity, Renal Comorbidity, Diabetes Comorbidity
Model 149	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Lung Comorbidity, Cardio Comorbidity, Renal Comorbidity, Smoking Comorbidity

Table A.5.18. Overview of COVID-19 prediction models along with quadruple combinations of the predictor variables – Part 7

Models	Predictor Variables
Model 150	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Lung Comorbidity, Cardio Comorbidity, Renal Comorbidity, Obesity Comorbidity
Model 151	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Lung Comorbidity, Cardio Comorbidity, Diabetes Comorbidity, Smoking Comorbidity
Model 152	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Lung Comorbidity, Cardio Comorbidity, Diabetes Comorbidity, Obesity Comorbidity
Model 153	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Lung Comorbidity, Cardio Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 154	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Lung Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Smoking Comorbidity
Model 155	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Lung Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Obesity Comorbidity
Model 156	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Lung Comorbidity, Renal Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 157	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Lung Comorbidity, Diabetes Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 158	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Cardio Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Smoking Comorbidity
Model 159	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Cardio Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Obesity Comorbidity

Table A.5.19. Overview of COVID-19 prediction models along with quadruple combinations of the predictor variables – Part 8

Models	Predictor Variables
Model 160	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Cardio Comorbidity, Renal Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 161	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Cardio Comorbidity, Diabetes Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 162	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Renal Comorbidity, Diabetes Comorbidity, Smoking Comorbidity, Obesity Comorbidity

A.5. Prediction Models Created Using Quintuple Combinations

By quintuple combinations of the predictor variables, a total of fifty-six prediction models have been formed, respectively. Table A.5.20 through Table A.5.25 show the overview of COVID-19 prediction models along with quintuple combinations of the predictor variables.

Table A.5.20. Overview of COVID-19 prediction models along with quintuple combinations of the predictor variables – Part 1

Models	Predictor Variables
Model 163	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Lung Comorbidity, Cardio Comorbidity, Renal Comorbidity
Model 164	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Lung Comorbidity, Cardio Comorbidity, Diabetes Comorbidity
Model 165	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Lung Comorbidity, Cardio Comorbidity, Smoking Comorbidity
Model 166	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Lung Comorbidity, Cardio Comorbidity, Obesity Comorbidity
Model 167	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Lung Comorbidity, Renal Comorbidity, Diabetes Comorbidity
Model 168	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Lung Comorbidity, Renal Comorbidity, Smoking Comorbidity
Model 169	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Lung Comorbidity, Renal Comorbidity, Obesity Comorbidity
Model 170	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Lung Comorbidity, Diabetes Comorbidity, Smoking Comorbidity

Table A.5.21. Overview of COVID-19 prediction models along with quintuple combinations of the predictor variables – Part 2

Models	Predictor Variables
Model 171	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Lung Comorbidity, Diabetes Comorbidity, Obesity Comorbidity
Model 172	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Lung Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 173	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Cardio Comorbidity, Renal Comorbidity, Diabetes Comorbidity
Model 174	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Cardio Comorbidity, Renal Comorbidity, Smoking Comorbidity
Model 175	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Cardio Comorbidity, Renal Comorbidity, Obesity Comorbidity
Model 176	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Cardio Comorbidity, Diabetes Comorbidity, Smoking Comorbidity
Model 177	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Cardio Comorbidity, Diabetes Comorbidity, Obesity Comorbidity
Model 178	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Cardio Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 179	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Renal Comorbidity, Diabetes Comorbidity, Smoking Comorbidity
Model 180	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Renal Comorbidity, Diabetes Comorbidity, Obesity Comorbidity

Table A.5.22. Overview of COVID-19 prediction models along with quintuple combinations of the predictor variables – Part 3

Models	**Predictor Variables**
Model 181	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Renal Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 182	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Diabetes Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 183	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Lung Comorbidity, Cardio Comorbidity, Renal Comorbidity, Diabetes Comorbidity
Model 184	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Lung Comorbidity, Cardio Comorbidity, Renal Comorbidity, Smoking Comorbidity
Model 185	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Lung Comorbidity, Cardio Comorbidity, Renal Comorbidity, Obesity Comorbidity
Model 186	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Lung Comorbidity, Cardio Comorbidity, Diabetes Comorbidity, Smoking Comorbidity
Model 187	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Lung Comorbidity, Cardio Comorbidity, Diabetes Comorbidity, Obesity Comorbidity
Model 188	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Lung Comorbidity, Cardio Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 189	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Lung Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Smoking Comorbidity
Model 190	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Lung Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Obesity Comorbidity

Table A.5.23. Overview of COVID-19 prediction models along with quintuple combinations of the predictor variables – Part 4

Models	Predictor Variables
Model 191	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Lung Comorbidity, Renal Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 192	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Lung Comorbidity, Diabetes Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 193	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Cardio Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Smoking Comorbidity
Model 194	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Cardio Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Obesity Comorbidity
Model 195	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Cardio Comorbidity, Renal Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 196	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Cardio Comorbidity, Diabetes Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 197	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Renal Comorbidity, Diabetes Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 198	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Lung Comorbidity, Cardio Comorbidity, Renal Comorbidity, Diabetes Comorbidity
Model 199	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Lung Comorbidity, Cardio Comorbidity, Renal Comorbidity, Smoking Comorbidity
Model 200	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Lung Comorbidity, Cardio Comorbidity, Renal Comorbidity, Obesity Comorbidity

Table A.5.24. Overview of COVID-19 prediction models along with quintuple combinations of the predictor variables – Part 5

Models	**Predictor Variables**
Model 201	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Lung Comorbidity, Cardio Comorbidity, Diabetes Comorbidity, Smoking Comorbidity
Model 202	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Lung Comorbidity, Cardio Comorbidity, Diabetes Comorbidity, Obesity Comorbidity
Model 203	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Lung Comorbidity, Cardio Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 204	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Lung Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Smoking Comorbidity
Model 205	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Lung Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Obesity Comorbidity
Model 206	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Lung Comorbidity, Renal Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 207	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Lung Comorbidity, Diabetes Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 208	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Cardio Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Smoking Comorbidity
Model 209	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Cardio Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Obesity Comorbidity
Model 210	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Cardio Comorbidity, Renal Comorbidity, Smoking Comorbidity, Obesity Comorbidity

Table A.5.25. Overview of COVID-19 prediction models along with quintuple combinations of the predictor variables – Part 6

Models	Predictor Variables
Model 211	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Cardio Comorbidity, Diabetes Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 212	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Renal Comorbidity, Diabetes Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 213	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Lung Comorbidity, Cardio Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Smoking Comorbidity
Model 214	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Lung Comorbidity, Cardio Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Obesity Comorbidity
Model 215	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Lung Comorbidity, Cardio Comorbidity, Renal Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 216	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Lung Comorbidity, Cardio Comorbidity, Diabetes Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 217	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Lung Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 218	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Cardio Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Smoking Comorbidity, Obesity Comorbidity

A.6. Prediction Models Created Using Sextuple Combinations

By using sextuple combinations of the predictor variables, a total of twenty-eight prediction models have been formed, respectively. Table A.5.26 through Table A.5.29 show the overview of COVID-19 prediction models along with sextuple combinations of the predictor variables.

Table A.5.26. Overview of COVID-19 prediction models along with sextuple combinations of the predictor variables – Part 1

Models	Predictor Variables
Model 219	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Lung Comorbidity, Cardio Comorbidity, Renal Comorbidity, Diabetes Comorbidity
Model 220	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Lung Comorbidity, Cardio Comorbidity, Renal Comorbidity, Smoking Comorbidity
Model 221	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Lung Comorbidity, Cardio Comorbidity, Renal Comorbidity, Obesity Comorbidity
Model 222	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Lung Comorbidity, Cardio Comorbidity, Diabetes Comorbidity, Smoking Comorbidity
Model 223	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Lung Comorbidity, Cardio Comorbidity, Diabetes Comorbidity, Obesity Comorbidity
Model 224	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Lung Comorbidity, Cardio Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 225	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Lung Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Smoking Comorbidity

Table A.5.27. Overview of COVID-19 prediction along with sextuple combinations of the predictor variables – Part 2

Models	Predictor Variables
Model 226	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Lung Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Obesity Comorbidity
Model 227	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Lung Comorbidity, Renal Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 228	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Lung Comorbidity, Diabetes Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 229	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Cardio Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Smoking Comorbidity
Model 230	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Cardio Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Obesity Comorbidity
Model 231	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Cardio Comorbidity, Renal Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 232	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Cardio Comorbidity, Diabetes Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 233	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Renal Comorbidity, Diabetes Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 234	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Lung Comorbidity, Cardio Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Smoking Comorbidity

Table A.5.28. Overview of COVID-19 prediction along with sextuple combinations of the predictor variables – Part 3

Models	Predictor Variables
Model 235	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Lung Comorbidity, Cardio Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Obesity Comorbidity
Model 236	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Lung Comorbidity, Cardio Comorbidity, Renal Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 237	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Lung Comorbidity, Cardio Comorbidity, Diabetes Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 238	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Lung Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 239	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Cardio Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 240	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Lung Comorbidity, Cardio Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Smoking Comorbidity
Model 241	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Lung Comorbidity, Cardio Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Obesity Comorbidity
Model 242	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Lung Comorbidity, Cardio Comorbidity, Renal Comorbidity, Smoking Comorbidity, Obesity Comorbidity

Table A.5.29. Overview of COVID-19 prediction along with sextuple combinations of the predictor variables – Part 4

Models	Predictor Variables
Model 243	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Lung Comorbidity, Cardio Comorbidity, Diabetes Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 244	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Lung Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 245	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Cardio Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 246	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Lung Comorbidity, Cardio Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Smoking Comorbidity, Obesity Comorbidity

A.7. Prediction Models Created Using Septuple Combinations

By using septuple combinations of the predictor variables, a total of eight prediction models have been formed, respectively. Table A.5.30 through Table A.5.31 show the overview of COVID-19 prediction models along with septuple predictor variables.

Table A.5.30. Overview of COVID-19 prediction models along with septuple combinations of the predictor variables – Part 1

Models	**Predictor Variables**
Model 247	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Lung Comorbidity, Cardio Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Smoking Comorbidity
Model 248	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Lung Comorbidity, Cardio Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Obesity Comorbidity
Model 249	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Lung Comorbidity, Cardio Comorbidity, Renal Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 250	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Lung Comorbidity, Cardio Comorbidity, Diabetes Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 251	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Lung Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 252	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Cardio Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Smoking Comorbidity, Obesity Comorbidity

Table A.5.31. Overview of COVID-19 prediction models along with septuple combinations of the predictor variables – Part 2

Models	Predictor Variables
Model 253	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Lung Comorbidity, Cardio Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Smoking Comorbidity, Obesity Comorbidity
Model 254	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Race, Lung Comorbidity, Cardio Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Smoking Comorbidity, Obesity Comorbidity

A.8. Prediction Models Created Using Octuple Combinations

By using octuple combinations of the predictor variables, a total of one prediction models have been formed, respectively. Table A.5.32 shows the overview of COVID-19 prediction models along with octuple predictor variables.

Table A.5.32. Overview of COVID-19 prediction models along with octuple combinations of the predictor variables

Models	Predictor Variables
Model 255	Age, Fever, Cough, Sex, Breathing Difficulty, Runny Nose, Throat Pain, Headache, Diarrhea, Pregnant, Race, Lung Comorbidity, Cardio Comorbidity, Renal Comorbidity, Diabetes Comorbidity, Smoking Comorbidity, Obesity Comorbidity

A.9. Feature Selection Based Prediction Models

Table A.5.33 and Table A.5.34 show the overview of Relief-F-based COVID-19 prediction models along with their predictor variables.

Table A.5.33. Overview of Relief-F-based prediction models for COVID-19 - Part 1

Models	Predictor Variables
Model 256	Age, race, fever, cough, headache, cardio comorbidity, runny nose, diarrhea, throat pain, breathing difficulty, sex, diabetes comorbidity, obesity comorbidity, pregnant, lung comorbidity, smoking comorbidity, renal comorbidity
Model 257	Age, race, fever, cough, headache, cardio comorbidity, runny nose, diarrhea, throat pain, breathing difficulty, sex, diabetes comorbidity, obesity comorbidity, pregnant, lung comorbidity, smoking comorbidity
Model 258	Age, race, fever, cough, headache, cardio comorbidity, runny nose, diarrhea, throat pain, breathing difficulty, sex, diabetes comorbidity, obesity comorbidity, pregnant, lung comorbidity
Model 259	Age, race, fever, cough, headache, cardio comorbidity, runny nose, diarrhea, throat pain, breathing difficulty, sex, diabetes comorbidity, obesity comorbidity, pregnant
Model 260	Age, race, fever, cough, headache, cardio comorbidity, runny nose, diarrhea, throat pain, breathing difficulty, sex, diabetes comorbidity, obesity comorbidity
Model 261	Age, race, fever, cough, headache, cardio comorbidity, runny nose, diarrhea, throat pain, breathing difficulty, sex, diabetes comorbidity
Model 262	Age, race, fever, cough, headache, cardio comorbidity, runny nose, diarrhea, throat pain, breathing difficulty, sex
Model 263	Age, race, fever, cough, headache, cardio comorbidity, runny nose, diarrhea, throat pain, breathing difficulty
Model 264	Age, race, fever, cough, headache, cardio comorbidity, runny nose, diarrhea, throat pain

Table A.5.34. Overview of Relief-F-based prediction models for COVID-19 – Part 2

Models	Predictor Variables
Model 265	Age, race, fever, cough, headache, cardio comorbidity, runny nose, diarrhea
Model 266	Age, race, fever, cough, headache, cardio comorbidity, runny nose
Model 267	Age, race, fever, cough, headache, cardio comorbidity
Model 268	Age, race, fever, cough, headache
Model 269	Age, race, fever, cough
Model 270	Age, race, fever
Model 271	Age, race
Model 272	Age

Table A.5.35 and Table A.5.36 show the overview of mRMR-based COVID-19 prediction models along with their predictor variables.

Table A.5.35. Overview of mRMR-based prediction models for COVID-19 – Part 1

Models	Predictor Variables
Model 273	Fever, cough, headache, age, breathing difficulty, obesity comorbidity, throat pain, diarrhea, runny nose, pregnant, smoking comorbidity, renal comorbidity, race, lung comorbidity, sex, diabetes comorbidity, cardio comorbidity
Model 274	Fever, cough, headache, age, breathing difficulty, obesity comorbidity, throat pain, diarrhea, runny nose, pregnant, smoking comorbidity, renal comorbidity, race, lung comorbidity, sex, diabetes comorbidity
Model 275	Fever, cough, headache, age, breathing difficulty, obesity comorbidity, throat pain, diarrhea, runny nose, pregnant, smoking comorbidity, renal comorbidity, race, lung comorbidity, sex
Model 276	Fever, cough, headache, age, breathing difficulty, obesity comorbidity, throat pain, diarrhea, runny nose, pregnant, smoking comorbidity, renal comorbidity, race, lung comorbidity
Model 277	Fever, cough, headache, age, breathing difficulty, obesity comorbidity, throat pain, diarrhea, runny nose, pregnant, smoking comorbidity, renal comorbidity, race
Model 278	Fever, cough, headache, age, breathing difficulty, obesity comorbidity, throat pain, diarrhea, runny nose, pregnant, smoking comorbidity, renal comorbidity
Model 279	Fever, cough, headache, age, breathing difficulty, obesity comorbidity, throat pain, diarrhea, runny nose, pregnant, smoking comorbidity
Model 280	Fever, cough, headache, age, breathing difficulty, obesity comorbidity, throat pain, diarrhea, runny nose, pregnant
Model 281	Fever, cough, headache, age, breathing difficulty, obesity comorbidity, throat pain, diarrhea, runny nose

Table A.5.36. Overview of mRMR-based prediction models for COVID-19 – Part 2

Models	Predictor Variables
Model 282	Fever, cough, headache, age, breathing difficulty, obesity comorbidity, throat pain, diarrhea
Model 283	Fever, cough, headache, age, breathing difficulty, obesity comorbidity, throat pain
Model 284	Fever, cough, headache, age, breathing difficulty, obesity comorbidity
Model 285	Fever, cough, headache, age, breathing difficulty
Model 286	Fever, cough, headache, age
Model 287	Fever, cough, headache
Model 288	Fever, cough
Model 289	Fever