CUKUROVA UNIVERSITY

INSTITUTE OF NATURAL AND APPLIED SCIENCES

MSc THESIS

Energy Consumption Forecasting: A Deep Learning Approach Based on A-LSTM Networks

Abdel Hack BIO BOULOU

Department of Satistics

July, 2025

CUKUROVA UNIVERSITY

INSTITUTE OF NATURAL AND APPLIED SCIENCES

MSc THESIS APPROVAL

Energy Consumption Forecasting: A Deep Learning Approach Based on A-LSTM Networks

---

Abdel Hack BIO BOULOU

Department of Statistics

This Master's Thesis was evaluated by the following Jury Members on 31/07/2025 and was approved by unanimity / ~~majority~~ of votes.

Jury : Prof. Dr. Mahmude Revan ÖZKALE ATICIOĞLU (Advisor) ……

: Prof. Dr. Mutlu AVCI ……

: Assoc. Prof. Dr. Murat GENÇ ……

This Thesis was written in the Department of Statistics, Institute of Natural and Applied Sciences.

Thesis Number:

Prof. Dr. Sadık DİNÇER
Director
Institute of Natural and Applied Sciences

Note: The usage of the presented specific declarations, tables, figures, and photographs either in this thesis or in any other reference without citation is subject to "The law of Arts and Intellectual Products" number of 5846 of Turkish Republic

CONTENTS

ABSTRACT I
ÖZ II
GENİŞLETİLMİŞ ÖZET III
ACKNOWLEDGEMENTS VII
LIST OF TABLES VIII
LIST OF FIGURES IX
SYMBOLS AND ABBREVIATIONS X
GENERAL INTRODUCTON 1
CHAPTER I: Fossil and Renewable Energies: Overview, Challenges, and Transition Policies 4
1.1. Introduction 4
1.2. Global state of fossil energies 4
1.3. Energy Transition 5
1.4 Renewable energy in an international context 6
1.4.1. Solar energy or photovoltaics (PV) 6
1.4.2. Wind Energy 6
1.4.3. Hydropower 8
1.4.4. Biomass Energy 9
1.4.5. Geothermal Energy 9
1.5. Fossil and renewable energy issues, policies and objectives 10
1.6. Conclusion 11
CHAPTER II: Artificial Neuronal Networks 14
2.1. Introduction 14
2.2. Why ANNs 15
2.3. Structure of artificial neural networks 16
2.4. Main architectures of ANNs 19
2.4.1. Feedforward Neural Networks (FNN) 19
2.4.2. Récurrent Neural Networks (RNN) 19
2.4.3. Convolutionnal Neural Networks (CNNs) 20
2.5. Neural Network Approaches to Time Series Forecasting 20
2.5.1. Auto-Regressive Integrated Moving Average (ARIMA) 20
2.5.2. ANN approach to time series modeling 22
2.5.2.1. Multi-layer Perceptron (MLP) 22
2.5.2.2. Long Short-Term Memory (LSTM) 23
2.5.2.2.(1) LSTM Classic 23
2.5.2.2.(2) Attention mechanism 26
2.5.2.2.(3) Attention-based LSTM (A-LSTM) Model 27
2.5.2.2.(4) Structure of the A-LSTM model 28

2.5.2.3. ARIMA-ANN model 30
2.5.2.4. The particularity of the A-LSTM model and what makes it an efficient model 31
2.6 Conclusion 32
CHAPTER III: Model Design and Tools 34
3.1. Introduction 34
3.2. General architecture of the proposed model 34
3.2.1. Data Preprocessing 35
3.2.1.1. Description of the database 35
3.2.1.2. Handling of Missing Values 39
3.2.1.3. Analysis and Processing of the Different Series 41
3.2.1.4. Data Normalization 46
3.2.2. Data Restructuring for Model Input 47
3.2.2.1. Creation of Time Series Sequences 47
3.2.2.2. Data Splitting 48
3.2.3. Training Process 48
3.2.4. Hyperparameters 49
3.2.5. Prediction 51
3.3. Implementation Phase 51
3.3.1. Development Environment 51
3.3.2. Libraries Used 52
3.3.3. Learning Parameters 52
3.4. Model evaluation metrics 53
3.5. Conclusion 54
CHAPTER IV: Results and Discussion 56
4.1. Introduction 56
4.2. Presentation of Training Results 56
4.3. Loss Curve 57
4.3.1. Loss Curve after renewable energy training 57
4.3.2. Loss curve after fossil energy training 58
4.4. Prediction 59
4.4.1. Prediction of renewable energy consumption 59
4.4.2. Prediction of fossil energy consumption 61
4.5. Comparison of the Model with Other Approaches 63
4.5.1. Description of the Models 63
4.5.2. Training outcome 64
4.6. General Discussion 66
4.6.1. Projection of Global Energy Dynamics by 2026 66
4.6.1.1. Renewable Energy 66
4.6.1.2. Fossil Energy 67

4.6.1.3. Limitations of Our Study 68
4.7. Conclusion 68
GENERAL CONLUSION 70
REFERECENCES 72
CURRICULUM VITAE 80

CUKUROVA UNIVERSITY

INSTITUTE OF NATURAL AND APPLIED SCIENCES

MSc THESIS

Energy Consumption Forecasting: A Deep Learning Approach Based on A-LSTM Networks

Abdel Hack BIO BOULOU

Advisor: Prof. Dr. Mahmude Revan ÖZKALE ATICIOĞLU

Department of Statistics

ABSTRACT

In the face of global energy instability, driven by the geopolitical crisis between Ukraine and Russia, persistent inflation, the depletion of fossil fuels, and the climate emergency, understanding and anticipating energy consumption trends has become a strategic priority. This thesis tackles this issue by developing a forecasting model based on artificial neural networks, more precisely an A-LSTM architecture (Attention-based Long Short-Term Memory), to predict short-term global consumption of fossil and renewable energies.

The study relies on historical data from covering the period 1965–2022 and including seven major world regions. Two datasets were constructed: one for fossil energy and the other for renewables. Each dataset contains seven input variables (regional consumption) and one output variable (global consumption). The data were analyzed and preprocessed using methods like normalization before being integrated into the model for training and prediction.

The proposed A-LSTM model comprises an LSTM layer, a Dropout layer to limit overfitting, an Attention layer to weigh past time steps, and a Dense output layer. Training was conducted in Google Colab using TensorFlow/Keras with the Adam optimizer and EarlyStopping.

The model demonstrated excellent generalization capabilities. It achieved respectively a Mean Absolute Error and Mean Absolute Percentage Error of 0,0002 and 1.77% for renewable energy and 0,0006 and 2.07% for fossil energy. This model has been then compared to some literature models like MLP, classical LSTM, ARIMA-ANN. The A-LSTM proposed consistently outperformed them, confirming the added value of attention mechanisms in enhancing time series forecasting. Forecasts to 2026 indicate continued growth in renewable energy consumption (average annual rate of 5.73%), while fossil energy remains stagnant (0.02%), suggesting a slow but steady global energy transition. this thesis demonstrates the high potential of the A-LSTM model in the field of energy forecasting, highlighting its accuracy, robustness, and adaptability; it offers an effective tool for policymakers and stakeholders in the energy sector.

Keywords: A-LSTM, Fossil, Renewable, energy, model.

ÇUKUROVA ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

Enerji Tüketimi Tahmini: A-LSTM Ağlarına Dayalı Derin Öğrenme Yaklaşımı

Abdel Hack BIO BOULOU

Danışman: Prof. Dr. Mahmude Revan ÖZKALE ATICIOĞLU

İstatistik Anabilim Dalı

ÖZ

Ukrayna ile Rusya arasındaki jeopolitik kriz, kalıcı enflasyon, fosil yakıtların tükenmesi ve iklim acil durumu gibi faktörlerin tetiklediği küresel enerji istikrarsızlığı karşısında, enerji tüketim eğilimlerini anlamak ve öngörmek stratejik bir öncelik hâline gelmiştir. Bu tez, bu sorunu ele almak amacıyla, yapay sinir ağlarına dayalı ve daha spesifik olarak dikkat mekanizması içeren bir A-LSTM (Attention-based Long Short-Term Memory) mimarisi kullanarak, fosil ve yenilenebilir enerji kaynaklarının kısa vadeli küresel tüketimini tahmin etmeye yönelik bir model geliştirmektedir.

Çalışma, 1965-2022 dönemini kapsayan ve yedi büyük dünya bölgesini içeren tarihsel verilere dayanmaktadır. Biri fosil enerji, diğeri yenilenebilir enerji için olmak üzere iki veri seti oluşturulmuştur. Her bir veri seti, yedi giriş değişkeni (bölgesel tüketim) ve bir çıkış değişkeni (küresel tüketim) içermektedir. Veriler, modele entegre edilmeden önce normalizasyon gibi yöntemler kullanılarak analiz edilmiş ve ön işleme tabi tutulmuştur.

Önerilen A-LSTM modeli; bir LSTM katmanı, aşırı öğrenmeyi önlemek için bir Dropout katmanı, geçmiş zaman adımlarına ağırlık vermek için bir Attention katmanı ve nihai küresel tahmini sağlayan bir Dense katmanından oluşmaktadır. Modelin eğitimi, TensorFlow/Keras kullanılarak Google Colab ortamında, Adam optimizasyon algoritması ve EarlyStopping yöntemiyle gerçekleştirilmiştir.

Model, mükemmel genelleme yetenekleri sergilemiştir. Yenilenebilir enerji için sırasıyla Ortalama Mutlak Hata 0,0002 ve Ortalama Mutlak Yüzde Hata %1,77; fosil enerji için ise 0,0006 ve %2,07 olarak elde edilmiştir. Bu model daha sonra MLP, klasik LSTM ve ARIMA-ANN gibi literatürde yer alan bazı modellerle karşılaştırılmıştır. Önerilen A-LSTM modeli bu modellerin her birinden daha iyi performans göstermiş ve dikkat mekanizmasının zaman serisi tahminlerinde sağladığı katma değeri doğrulamıştır.

2026 yılına kadar yapılan öngörüler, yenilenebilir enerji tüketiminde %5,73'lük yıllık ortalama büyüme oranıyla sürekli bir artış olduğunu; fosil enerji tüketiminin ise %0,02 ile durağan kaldığını göstermektedir. Bu durum, küresel enerji karmasında yavaş ama istikrarlı bir geçişe işaret etmektedir. Bu tez, A-LSTM modelinin enerji tahminleme alanındaki yüksek potansiyelini ortaya koymakta; doğruluğu, sağlamlığı ve uyarlanabilirliği vurgulamakta ve enerji sektöründeki politika yapıcılar ile paydaşlar için etkili bir araç sunmaktadır**.**

Anahtar kelimeler: A-LSTM, Fosil, Yenilenebilir, enerji, model.

GENİŞLETİLMİŞ ÖZET

Genel Bağlam ve Konunun Gerekçesi

Mal ve hizmetlerin üretimi ve pazarlanmasından, dünyanın herhangi bir noktasına yapılan ulaşımlara, sağlık hizmetlerinden modern yaşamın günlük ihtiyaçlarına kadar uzanan geniş bir yelpazede insan, enerjinin yaşamının ayrılmaz bir parçası olduğu bir sistem içinde yer almaktadır. Enerjinin farklı alanlardaki kullanımları, onu ekosistem değerinin tahmininde yaygın olarak benimsenmiş bir ölçü birimi hâline getirmiştir (Jørgensen, 2006). Bu kaynak o kadar gereklidir ki, azalması dünya ekonomileri üzerinde ciddi sonuçlara ve toplumların yaşamında önemli aksaklıklara yol açabilir (Yasmeen ve Shah, 2024). Günümüzde bu enerji kaynaklarının yaklaşık %80'i fosil kaynaklardan sağlanmaktadır. Dolayısıyla, tedarik zincirinde meydana gelebilecek herhangi bir aksama; ekonomik büyümede yavaşlama, enflasyonun artması ve alım gücünün düşmesi gibi ciddi sonuçlara yol açabilir.

Nitekim, COVID-19 pandemisinin ardından 2022 yılında Rusya'nın Ukrayna'yı işgaliyle birlikte dünya, tarihinde ilk kez ciddi bir enerji kriziyle karşı karşıya kalmıştır (IEA, 2022a; Farghali ve diğ., 2023; Romei ve Smith, 2023). Bu çatışma, fosil yakıtların tedarik zincirini kesintiye uğratmış ve ülkelerin ekonomileri üzerinde derin etkiler yaratmıştır. Fosil enerji fiyatlarındaki enflasyon, dolaylı olarak mal ve hizmet fiyatlarını da artırmıştır. Hanehalkı enerji maliyetlerinin %62,6 ile %112 arasında arttığı ve bu durumun hane harcamalarında %2,7 ile %4,8 oranında artışa yol açtığı belirtilmektedir (Guan ve ark., 2023). Kontrol altına alınmakta zorlanılan bu enflasyon, yaşam maliyetlerinde keskin bir artışa neden olarak milyonlarca insanı aşırı yoksulluk sınırına itebilir (Farghali ve diğ., 2023). Pek çok ülkede enflasyondaki bu hızlanma, faiz oranlarının artırılmasına ve böylece ekonomik büyümenin yavaşlamasına neden olmuştur (IEA, 2022). Buna ek olarak, bu enerji türlerinin kullanımı iklim değişikliğine de katkıda bulunmaktadır. Fosil enerji tüketimi ile CO_2 emisyonları arasında güçlü bir korelasyon olduğu bilinmektedir; enerji talebinin tarihsel olarak en yüksek seviyelere ulaşmasıyla birlikte, CO_2 yoğunluğu da rekor düzeye ulaşmıştır. Dünya Meteoroloji Örgütü'ne (WMO) göre, 2024 yılı tarihte kaydedilen en sıcak yıl olmuştur (WMO, 2025). Zaten sanayi devriminden bu yana atmosferdeki CO_2 oranının %47 oranında arttığı bildirilmektedir (Aleklett, 2007; Abas ve ark., 2015'ten aktaran).

Fosil kaynakların jeopolitik istikrarsızlık ve iklim krizi nedeniyle giderek azaldığı bir dönemde, dünya genelinde enerjiyle ilgili endişeler ve enerji dönüşümüne dair tartışmalar daha önce hiç olmadığı kadar yoğunlaşmıştır. 2030 yılına kadar sürdürülebilir bir enerji geçişi hedefleyen Paris Anlaşması çerçevesinde, bu dönüşümün gerçekleşebilmesi için tüketim dinamiklerinin derinlemesine anlaşılması hayati önemdedir. Bu bağlamda, küresel enerji tüketiminin (özellikle fosil ve yenilenebilir kaynakların) gelecekteki eğilimlerini analiz etmek ve öngörmek, hem hükümetler hem de enerji sektörü aktörleri için stratejik bir zorunluluk hâline gelmiştir.

Ancak geleneksel tahmin yöntemleri genellikle doğrusal varsayımlara ve klasik istatistiksel modellere dayanmakta olup, enerji zaman serilerinin karmaşıklığını ve doğrusal olmayan yapısını yeterince yakalayamamaktadır. Bu sınırlamalara karşılık, derin öğrenme (deep learning) yöntemleri ve özellikle yinelemeli sinir ağları (recurrent neural networks), umut vadeden alternatif araçlar olarak öne çıkmaktadır. İşte bu bağlamda, bu tez çalışması, 1965-2022 dönemine ait tarihsel verileri kullanarak, küresel enerji tüketimini tahmin etmeyi amaçlayan A-LSTM (dikkat mekanizmalı uzun-kısa vadeli bellek ağı) temelli bir öngörü modelinin geliştirilmesini ve değerlendirilmesini hedeflemektedir.

Araştırmanın Amaçları

Bu araştırmanın temel amacı, dikkat mekanizmasıyla güçlendirilmiş Uzun Kısa Süreli Bellek (A-LSTM) ağı kullanarak, yenilenebilir ve fosil enerji kaynaklarının küresel tüketimindeki gelecekteki eğilimleri modellemek ve tahmin etmektir. Bu bağlamda, 2030 yılına kadar yenilenebilir enerji kaynaklarının fosil yakıtların yerini alıp alamayacağını değerlendirmek hedeflenmektedir.
Araştırmanın özel hedefleri şunlardır:

- Yenilenebilir ve fosil enerji tüketimini doğru bir şekilde tahmin edebilecek ve karmaşık zamansal dinamikleri yakalayabilecek nitelikte optimal bir A-LSTM yapılandırmasının belirlenmesi,
- Önerilen A-LSTM modelinin performansının, diğer modellerle karşılaştırılarak değerlendirilmesi,
- Geliştirilen model aracılığıyla incelenen enerji türlerinin tüketim eğilimlerinin dinamiklerinin analiz edilmesi.

Yöntem

Bu tez kapsamında, 1965-2022 dönemi boyunca fosil ve yenilenebilir enerji tüketimini ayrı ayrı tahmin etmek amacıyla dikkat mekanizmasına sahip bir A-LSTM (Long Short-Term Memory) modeline dayalı bir yöntem önerilmiştir. Bu yaklaşım, her iki enerji türünün seyrini incelemeyi ve kısa ve orta vadede fosil yakıtların yenilenebilir enerjilerle ikame edilme potansiyelini değerlendirmeyi amaçlamaktadır.

Kullanılan veriler, dünya enerji istatistiklerinin kalitesiyle tanınan Energy Institute (EI) veri tabanından alınmıştır. Bu veriler 58 yıllık bir dönemi (1965-2022) kapsamaktadır ve yedi büyük bölgeyi içermektedir: Kuzey Amerika, Güney Amerika, Avrupa, BDT (CIS), Orta Doğu, Afrika ve Asya-Pasifik. Her veri tabanı (fosil veya yenilenebilir) sekiz değişken içermektedir: yedisi bölgelere karşılık gelen açıklayıcı değişkenler ve sonuncusu hedef değişken olarak kullanılan dünya enerji tüketimini temsil eden değişkendir. Veri tabanlarının oluşturulması, ulusal enerji tüketimlerinin

bölgesel düzeyde toplanmasını ve zaman uyumunun sağlanmasını gerektirmiştir. Bu aşamada eksik değerlerin dikkatli bir şekilde işlenmesi yapılmıştır.

Serilerin görsel analizi, güçlü bir trend bileşeni ve doğrusal olmayan bir yapı olduğunu, ancak belirgin bir mevsimsellik bulunmadığını ortaya koymuştur. Bu gözlem, tüm serilerin durağan olmadığını gösteren ADF (Augmented Dickey-Fuller) testleriyle doğrulanmıştır. Ancak, sistematik olarak fark almaya başvurmaktansa, serilerin ham hâlleriyle korunmasına karar verilmiştir. Bu karar, LSTM türü ağların durağan olmayan verileri işleyebilme ve altta yatan dinamikleri (trend, döngüler, doğrusal olmayan bağımlılıklar) otomatik olarak yakalayabilme yeteneğini ortaya koyan son literatüre dayanmaktadır. Öte yandan, öğrenme sırasında istikrarlı bir yakınsama sağlamak amacıyla, modelde tanh ve sigmoid gibi hassas aktivasyonların kullanılması nedeniyle tüm değişkenlere *Min-Max(.)* normalizasyonu uygulanmıştır.

Bir sonraki aşama, verilerin ardışık dört yıllık bir pencereye (time step = 4) göre kayan zaman dizilerine yeniden yapılandırılmasından oluşmuştur. Bu, her bir gözlem için modelin giriş olarak önceki dört yılın bölgesel enerji tüketimlerini alarak bir sonraki yılın dünya tüketimini tahmin etmesi anlamına gelmektedir. Mevcut 58 yıl üzerinden bu yöntemle 54 kullanılabilir dizi elde edilmiştir. Daha sonra veriler %80-%20 oranında ayrılmış; yani 43 tanesi eğitim (80 %) ve 11 tanesi test (20 %) için kullanılmıştır. Bu ayrımda zaman sıralaması korunmuştur.

Önerilen Modelin Mimarisi: A-LSTM

Uygulanan model aşağıdaki katmanlardan oluşmaktadır:

- LSTM Katmanı: Bölgesel değişkenler arasındaki karmaşık zamansal ilişkileri yakalamak için,
- Dropout Katmanı: Aşırı öğrenmeyi önlemek için,
- Dikkat (Attention) Katmanı: Geçmiş gözlemlerin göreli önemini ağırlıklandırmak için. Bu katman, LSTM çıkışındaki her zaman adımına bir skor atar, bu skorları softmax fonksiyonu aracılığıyla normalize edilmiş ağırlıklara dönüştürür ve ardından en anlamlı bilgiyi özetleyen bir bağlam vektörü üretir,
- Yoğun Katman (Çıkış Katmanı): Küresel enerji tüketiminin nihai tahminini üretmek için.

Modelin eğitimi, *Adam* algoritmasıyla gerçekleştirilmiştir.

Eğitime erken durdurma (*EarlyStopping*) fonksiyonları eşlik etmiş, yenilenebilir veri seti için 100 epoch, fosil veri seti için ise 200 epoch yapılandırması kullanılmıştır. Geliştirme ortamı olarak Google Colab üzerinde Python (3.10 versiyonu) tercih edilmiş; TensorFlow/Keras, scikit-learn, NumPy ve Matplotlib kütüphaneleri kullanılmıştır.

Modelin performansı, fosil ve yenilenebilir veriler üzerinde ayrı ayrı uygulanan dört klasik metrik ile değerlendirilmiştir: Ortalama Kare Hata (MSE), Karekök Ortalama Kare Hata (RMSE),

Ortalama Mutlak Hata (MAE) ve Ortalama Mutlak Yüzde Hata (MAPE). Bu yaklaşım, modelin yalnızca doğruluğunu değil, aynı zamanda iki farklı kaynak türü arasındaki enerji davranışlarını dikkate alarak yeni verilere genelleme yapabilme yeteneğini de ölçmeye olanak tanımaktadır.

Sonuçlar

Bu tez, 1965-2022 dönemini kapsayan çok değişkenli zaman serilerine uygulanan dikkat mekanizmasına sahip A-LSTM (Long Short-Term Memory) modeli aracılığıyla, küresel fosil ve yenilenebilir enerji tüketiminin kestirimsel modellenmesini incelemiştir. Temel amaç, önerilen modelin küresel enerji eğilimlerini doğru bir şekilde tahmin etme yeteneğini ortaya koymak ve bunu MLP, klasik LSTM ve hibrit ARIMA-ANN gibi referans mimarilerle karşılaştırmaktır.

Elde edilen sonuçlar, A-LSTM modelinin serilerin ardındaki dinamikleri mükemmel şekilde öğrendiğini göstermektedir. Yenilenebilir enerji verileri üzerinde, kayıp eğrisi ilk epoch'lardan itibaren düşük ve istikrarlı kalmış, 80. epoch'a kadar düzenli bir yakınsama göstermiştir ve aşırı öğrenme belirtisi gözlenmemiştir. Model, artan eğilimleri yüksek doğrulukla yakalamayı başarmış, 0,0002 MSE, 0,0116 MAE ve yalnızca %1,77 MAPE ile çok yüksek bir genelleme kalitesi sergilemiştir.

Fosil enerji verileri, daha karmaşık ve dengesiz olmasına rağmen, A-LSTM modeli güçlü bir performans sürdürmüştür. Model hızlı bir şekilde yakınsamış, en iyi performansını 10. epoch'ta göstermiştir. Tahmin edilen değerler, gerçek değerlere çok yakın seyretmiş; MAE 0,019 ; MAPE %2,07 ve RMSE 0,025 olarak kaydedilmiştir. 2020 yılındaki daha belirgin sapmalar pandemiye bağlanmış, bu da dışsal şokların öngörülmesindeki zorluğu teyit etmiştir.

Diğer modellerle karşılaştırıldığında, A-LSTM en başarılı model olarak öne çıkmıştır. MLP modeli önemli hatalar göstermiştir (yenilenebilir verilerde %12,76 MAPE), ARIMA-ANN modeli ise orta düzeyde sonuçlar elde etmiştir. Klasik LSTM modeli iyi bir performans sergilemiştir (yenilenebilir verilerde %2,66, fosil verilerde %2,31 MAPE), ancak dikkat mekanizmasından yararlanan A-LSTM'nin gerisinde kalmıştır. Bu farklılıklar, A-LSTM modelinin yalnızca sağlam değil, aynı zamanda enerji zaman serilerindeki karmaşık dinamiklere daha duyarlı olduğunu ortaya koymaktadır.

Son olarak, 2026 yılına kadar öngörüler üretilmiştir. Yenilenebilir enerjilerin yıllık ortalama %5,73 oranında güçlü bir büyüme sergilemesi beklenirken, fosil enerji tüketiminin yıllık ortalama %0,02 ile durağan bir seyir izleyeceği öngörülmektedir. Bu zıtlık, enerji geçişinin başlangıcına işaret ederken, mutlak hacim açısından fosil kaynakların hâlâ baskın olduğunu da göstermektedir.

Özetle, bu tez A-LSTM modelinin enerji öngörüleri alanındaki geçerliliğini doğrulamış, doğruluk, sağlamlık ve uyarlanabilirliği bir arada sunmuştur. Karmaşık verilerin öngörüsel analizine dair umut verici perspektifler açmakta ve gelecekte dışsal faktörlerin entegrasyonu ile daha gelişmiş mimarilerin araştırılması için bir temel oluşturmaktadır.

ACKNOWLEDGEMENTS

The completion of this work would not have been possible without the support of several people to whom I would like to express my deepest gratitude.

First and foremost, I extend my heartfelt thanks to Prof. Dr. Mahmude Revan ÖZKALE ATICIOĞLU, my thesis supervisor, for her invaluable and continuous guidance throughout this research. She was the one who introduced me to the fascinating world of neural networks, an intellectual journey that has marked a true turning point in my understanding of statistics and advanced data analysis. Her availability and kindness have touched me deeply, and I truly feel fortunate to have met her.

I would also like to express my sincere appreciation to my family, whose unwavering emotional support has accompanied me despite the physical distance. Thank you, Mom and Dad, for your love, prayers, and constant encouragement: I am endlessly grateful to you.

My thanks also go to the Department of Statistics and all its faculty members for the quality of their teaching and the solid, valuable knowledge they have shared throughout this master's program.

I am also grateful to the members of the jury for agreeing to evaluate this work and for their constructive feedback.

Finally, I would like to thank my friends Kemal CHABI, Julia IGNATOVA, Nassirou BASSIROU, as well as all those who, directly or indirectly, thought of me, supported me, or encouraged me throughout this journey.

LIST OF TABLES

Table 1: Some activation functions and their derivatives. 18
Table 2: List of countries considered 37
Table 3: A summary of each dataset used in the modeling process 39
Table 4: ADF results 44
Table 5: Order of differencing leading to stationarity. 45
Table 6: Optimized hyparameters for the A-LSTM model 51
Table 7: Absolue Errors between actual and predicted of renewable values. 60
Table 8: Metrics results_Renewable 61
Table 9: Absolue Errors between actual and predicted of fossil values. 62
Table 10: Metric results_Fossile 63
Table 11: Metrics results for different models using renewable datasets. 65
Table 12: Metrics results for different models using fossil dataset. 65

LIST OF FIGURES

Figure 1: The biological neuron structure (Neves et al., 2018).....14
Figure 2: Representation of an artificial neural network with two hidden layers.....16
Figure 3: Structure of an ANN.....17
Figure 4: LSTM Bloc.....24
Figure 5: Attention Mechanism steps.....27
Figure 6: An illustration of A-LSTM Bloc (Kumar et al., 2023).....30
Figure 7: Architecture of our proposed model A-LSTM.....35
Figure 8: Annual trends in renewable energy consumption by region from 1965-2022.....41
Figure 9: Ranking of Regions by Compound Annual Growth Rate.....42
Figure 10: Annual trends in fossil energy consumption by region from 1965-2022.....43
Figure 11: Ranking of Regions by Average Annual Growth Rate.....43
Figure 12: Overview of the data set after normalization.....47
Figure 13: A-LSTM network architecture implemented in Python.....49
Figure 14: Summary of A-LSTM parameters after data learning on Renewable data.....56
Figure 15: Summary of A-LSTM parameters after data learning on Fossile data.....57
Figure 16: Loss Curve of A-LSTM applied on Renewable data.....58
Figure 17: Loss Curve of A-LSTM applied on Fossile data.....59
Figure 18: Evolution of actual vs. predicted renewable data.....60
Figure 19: Evolution of actual vs. predicted Fossil data.....62
Figure 20: Comparative gloabal consumption forcasts (2022-2026) for fossil fuels vs renewables energies.....67

SYMBOLS AND ABBREVIATIONS

REN21	:	Renewable Energy Policy Network for the 21st Century
PCA	:	Paris Climate Agreement
SDG	:	Sustainable Development Goals
UN	:	United Nations
IEA	:	International Energy Agency
ANN	:	Artificial Neural Network
BCF	:	Billion Cubic Feet
MT	:	Million Tonnes
CO2	:	Carbon dioxide
NASA	:	National Aeronautics and Space Administration
IPCC	:	Intergovernmental Panel on Climate Change
WMO	:	World Meteorological Organization
MW	:	Megawatt
PV	:	Photovoltaic
GW	:	Gigawatt
LCOE	:	Levelized Cost of Energy
USD	:	United States Currency
IRENA	:	International Renewable Energy Agency
EJ	:	Exajoule
HP	:	Heat Pump
CO	:	Carbon Monoxide
COP11	:	The 11th Conference of the Parties
COP	:	Conference of the Parties
IMF	:	International Monetary Fund
MLP	:	Multi-Layer Perceptron
ARIMA	:	AutoRegressive Integrated Moving Average
CNN	:	Convolutional Neural Network
RNN	:	Recurrent Neural Network
FNN	:	Forward Neural Network
RBF	:	Radial Basis Function
AR	:	AutoRegressive
MA	:	Moving Average
ACF	:	Autocorrelation Function
PACF	:	Partial Autocorrelation Function
ADF	:	Augmented Dickey-Fuller test

DNN	:	Deep Neural Network
RBFN	:	Radial Basis Function Network
GRU	:	Gated Recurrent Unit
BP	:	BackPropagation
MSE	:	Mean Squared Error
MAE	:	Mean Absolute Error
RMSE	:	Root Mean Squared Error
MAPE	:	Mean Absolute Percentage Error
EI	:	Energy Institute
CIS	:	Commonwealth of Independent States
USSR	:	Union of Soviet Socialist Republics
CAGR	:	Compound Annual Growth Rate
API	:	Application Programming Interface
SARIMA	:	Seasonal AutoRegressive Integrated Moving Average

GENERAL INTRODUCTON

From the production and commercialization of goods and services to global transportation, healthcare systems, and modern daily life support, energy is an essential and inseparable component of human existence. Its diverse applications have made it a widely recognized metric in the estimation of ecosystem value (Jørgensen, 2006). This resource is so important that its scarcity could have profound consequences on global economies and on the lives of populations (Yasmeen and Shah, 2024). Currently, 80% of the world's energy resources are derived from fossil sources, namely oil, coal, and natural gas (IEA, 2024). Therefore, any disruption in the supply chain is likely to trigger consequences such as economic slowdown, inflation, and a decline in purchasing power.

This was notably observed following Russia's invasion of Ukraine in 2022, after the COVID-19 pandemic, as the world faced a true energy crisis (IEA, 2022a; Farghali et al., 2023; Romei and Smith, 2023). The conflict severely disrupted fossil fuel supply chains and had major repercussions, particularly on European economies (Vinuales, 2023). The surge in fossil fuel prices indirectly increased the costs of goods and services. Household energy costs rose by 62,6% to 112%, leading to an increase of 2,7% to 4.8% in household spending (Guan et al., 2023). Inflation remains difficult to stabilize, and the soaring cost of living risks pushing hundreds of millions of people into extreme poverty (Farghali et al., 2023). At the macroeconomic level, global economies were also affected. The energy crisis caused a rise in transportation costs, which in turn increased the prices of goods and services, reducing consumers' purchasing power (Farghali et al., 2023). The industrial sector faced reduced production, lower profit margins, and higher unemployment rates. In many countries, accelerating inflation led to rising interest rates, thereby slowing economic growth (IEA, 2022). This alarming situation has exacerbated existing uncertainty surrounding fossil energy, which is increasingly struggling to meet global demand. Moreover, the climate change effects linked to the use of fossil fuels are becoming more apparent. Rising CO2 levels are closely associated with fossil fuel consumption, and with global energy demand at an all-time high, CO2 emissions have reached unprecedented levels (Aleklett, 2007 as cited in Abas et al., 2015). The World Meteorological Organization (WMO) estimates that 2024 is the warmest year ever recorded. CO2 levels have already increased by 47% since the Industrial Revolution (NASA, 2017; IPCC, 2019).

In this context of growing global energy tensions, marked by the depletion of fossil resources due to recent geopolitical instability and the urgency of the climate crisis, concerns around energy consumption and the transition to sustainable sources have never been more pressing. The objective set by many nations to achieve energy transition and adopt viable

renewable energy sources by 2030, in line with the Paris Agreement (COP21), requires a precise understanding of consumption dynamics to anticipate future demand.

It is within this context that we have questioned the actual impact of renewable energy on the production and consumption of fossil fuels. Given their current expansion trend, can renewables become a central pillar of the global energy mix by 2030, overtaking fossil fuels in line with international energy transition goals?

Understanding and anticipating future trajectories of global energy consumption, especially for fossil and renewable sources, has thus become a major strategic priority for both governments and energy sector stakeholders.

Traditional forecasting methods, often based on linear assumptions or classical statistical models such as ARIMA and SARIMA, struggle to capture the complexity and non-linearity inherent in time series data, particularly in the energy domain (Wang et al., 2021; Tarmanini et al., 2023). In response to these limitations, deep learning techniques, especially recurrent neural networks (RNNs), have emerged as promising alternatives.

This research falls within that perspective and aims to develop and evaluate a predictive model (A-LSTM), based on a variant of recurrent neural networks, namely Long Short-Term Memory (LSTM), enhanced with an attention mechanism. This model is designed to forecast global energy consumption using historical data spanning from 1965 to 2022.

This document is structured as follows. Chapter 1 introduces the theoretical framework on energy. Chapter 2 presents the state of the art on artificial neural networks and details the A-LSTM model developed. Chapter 3 focuses on the model design and all the tools used in the process. Particular emphasis is placed on the configuration of network parameters to develop a precise prediction algorithm. Chapter 4 is dedicated to presenting the results after training the model on energy data. This includes performance evaluation and comparisons with other models from the literature to assess the model's robustness. Predictions are then carried out, and the impact of renewable energy on fossil energy is analyzed. Finally, the thesis concludes with a general summary of findings and perspectives.

Research objectives

The aim of this research is to model and forecast the future trajectory of global consumption of both renewable and fossil energies, using a Long Short-Term Memory network enhanced with an attention mechanism (A-LSTM). The aim is to assess whether renewable energy sources are on track to progressively replace fossil fuels by the year 2030.

Specifically, it will:

- Identify the optimal configuration of an A-LSTM capable of capturing the complex temporal dynamics and accurately predicting the consumption of renewable and fossil energies.
- Evaluate the performance of the proposed A-LSTM model compared to other models.
- Analyze the dynamics of consumption trends for the various energies studied, using the proposed model.

CHAPTER I: Fossil and Renewable Energies: Overview, Challenges, and Transition Policies

1.1 Introduction

Energy is a fundamental pillar of the economic and social development of nations. Since the Industrial Revolution, the world has relied primarily on fossil fuels, notably oil, natural gas, and coal. These resources, long regarded as essential drivers of growth, have helped meet an ever-increasing energy demand. Even today, they occupy a central place in the global energy mix, accounting for nearly 80% of primary consumption (IEA, 2024). However, this dependency raises many concerns due to their finite nature and environmental impact.

In this context, renewable energies have emerged as a credible alternative. Derived from inexhaustible natural sources such as the sun, wind, or water, they allow for energy production without harming the environment or compromising future generations. Although their share in global energy production remains modest, their development is now at the heart of energy transition strategies.

This chapter aims to provide a global overview of fossil and renewable energies, highlighting their respective challenges, recent developments, and the prospects they offer in a world striving for sustainability.

1.2 Global state of fossil energies

Fossil fuels dominate the energy sector. According to recent data, they represent about 87% of the primary energy used worldwide (Lyman, 2016). The shares of oil, gas, and coal, which make up fossil fuels, are estimated at approximately 36%, 27%, and 23%, respectively. Global production of oil, natural gas, and coal occurs at rates of 87,4 Mb (11,92 Mtoe), 329 BCF (8,225 Mtoe), and 21,63 MT (14.42 Mtoe) per day (Abas et al., 2015). Oil consumption continues to increase every year and is currently estimated at 103 million barrels per day (Mb/d). Predictions by the International Energy Agency (IEA, 2021) show that this consumption will rise to 105 Mb/day by 2028. The question of when the peak of oil production and consumption will occur continues to generate much debate, but what is certain, according to the authors, is that this peak will happen in a near future, or more precisely within the next decade (Andrews & Udall, 2003; IEA, 2013; Tea, 2008; Towler, 2014). Once this peak is reached, global reserves and production will begin to decline, likely indicating the actual decline profile. The finite nature of fossil fuels, therefore, shows the need to find alternatives to ensure continued energy security. Added to this are their negative environmental impacts. Indeed, it is evident that fossil fuels have significant implications in terms of greenhouse

gas emissions, climate change, and environmental sustainability (Gustavsson et al., 2021; Wang et al., 2022; Chakhtoura, 2023; Grand and Candel, 2023). The CO2 level has increased by 47% since the Industrial Revolution, raising the Earth's temperature by 1.9 degrees Fahrenheit and sea levels by 7 inches (NASA, 2017; IPCC, 2019). World Meteorological Organization (WMO) specified that 2024 was the warmest year on record at about 1,5ºC (WMO, 2025). According to some ecologists, the peak in oil production and consumption leads to a peak in CO2 emissions in the atmosphere, which is also continually increasing (402 ppm) (Aleklett, 2007 cited by Abas et al., 2015).

The limitation of fossil fuel reserves and the alarming environmental consequences caused by their use have led to the necessity of finding alternative energy sources and promoting policies aimed at reducing dependency on these energy sources (Vivien, 2004). This gave rise to renewable, or "green," energy, recognized as much cleaner, unlimited, and non-damaging to the environment. This shift from fossil fuels to renewable energy is known as the "Energy Transition."

1.3 Energy Transition

Energy is one of the most important raw materials in modern society, powering daily life, economies, and transportation. In 2020, global energy consumption was 30,01% for industry, 26,18% for transportation, and 22,08% for the residential sector. As we mentioned earlier, the two most widely used forms of energy today are fossil fuels and renewable energy. According to estimates by British Ecotricity, annual fossil fuel consumption is 11 billion tons (Kalair et al., 2021). Even if we maintain the current pace, the vision centered on the use of renewable energy is based on significantly reducing dependence on fossil fuels, whose quantity is limited and which are the main contributors to climate change and greenhouse gas emissions (Roos, 2024).

Renewable energies are forms of energy derived from natural sources such as the sun, wind, water flows, and the Earth's natural heat, which can produce final forms of energy (electricity, heat, or fuel). One of the most important aspects emphasized by everyone is their unlimited nature, as they come from inexhaustible natural sources (sun, heat, water) and have no impact on the environment, unlike conventional energy sources that rely on limited fossil fuel reserves (coal, gas, oil) (Collard, 2015). Today, the share of renewable energy in the global energy mix is estimated to be around 20% (Kalair et al., 2021). This figure remains low considering the goal of achieving net-zero carbon emissions by 2050 (IEA, 2021). According to Wang et al. (2022), the total capacity of renewable energy increased from 1,223,533 MW in 2010 to 2,532,866 MW in 2019, and more than 80% of electricity could come from renewables by the middle of the century.

Overall, we distinguish six major forms of renewable energy: solar energy, wind energy, hydropower, biomass, geothermal energy, and tidal or wave energy (Voyant, 2011)

1.4 Renewable energy in an international context

1.4.1 Solar energy or photovoltaics (PV)

Photovoltaic energy is one of the renewable and clean energy sources that do not harm the environment. It (PV) is produced by directly converting photons from light radiation (solar or other) into electricity. It uses photovoltaic modules made up of cells or photoelectric cells, manufactured from materials sensitive to visible wavelengths, which perform this energy transformation (Demraoui et al., 2023). This process does not require any intermediate thermodynamic cycle, meaning the radiation is directly converted into electricity without the intermediate use of heat (unlike thermodynamic solar energy). Photovoltaic cells harness the photoelectric effect to produce direct current by absorbing solar radiation. This effect allows the cells to directly convert the light energy from photons into electricity through a semiconductor material that carries electric charges. Thus, an electric current is generated from solar radiation. Since the sun is the main source of this energy, it is an inexhaustible resource that can be used anywhere in the world (Voyant, 2011). This energy is generally used for electricity production and for heating water and household spaces.

In 2021, PVs generated about 3,7% of global electricity. Although this rate remains low, it is important to note that increasing efforts are being made to boost the production and consumption of this energy, especially in the electricity sector. In 2021, approximately 843 gigawatts (GW) of installed solar photovoltaic systems were estimated globally, which is 21 times more than what was installed in 2010 (Osman et al., 2023). Between 2020 and 2021, there was a 13% increase in installed photovoltaic systems (a total of 133 GW), representing the highest increase of all renewable sources in 2021 (IRENAa, 2022). This growth is largely supported by technological advancements and policies promoting the consumption of this green energy. According to the IEA (2023) report, PV costs dropped significantly by 88,89% between 2010 and 2022. The global average LCOE (Levelized Cost of Electricity) decreased from 0,45 USD/kWh in 2010 to 0,05 USD/kWh in 2022. In the past, this energy was not accessible to everyone due to its high manufacturing costs.

The current momentum related to the production and accessibility of PVs is still modest in light of the 2050 decarbonization goal, where renewable energies are expected to cover 90% of the world's electricity supply. Its combination with wind energy should represent over 52% of electricity sector sources. Speaking of wind energy, what is its current situation?

1.4.2 Wind Energy

Wind energy is by far one of the oldest renewable energies and the most widely used after hydropower. Its first uses date back to ancient times when "windmills" converted wind energy into mechanical energy to grind grain or used wind turbines coupled with water pumps to irrigate arid

areas for agriculture and provide drinking water for animals. The advent of electricity enabled the association of wind turbines with a generator (wind generator) (Kose et al., 2004; Ferkous, 2009).

This wind generator allowed for the direct conversion of wind into electricity. To go into more detail, the mechanism for producing electricity from wind sources is based on converting the kinetic energy created by the wind into mechanical energy, and then into electricity. A wind turbine consists of a mast, a rotor, a generator, and several blades, all connected to a transformer. This mechanical energy is then transformed into electricity via a generator located in the nacelle. The electrical energy thus converted can be directly injected into an electrical grid, but it can also be stored in various ways, similar to photovoltaic energy (batteries, pumped storage stations, flywheels, etc.) (Johnson, 1985; Blaabjerg & Ma, 2017; Wagner, 2020). The constant development of powerful electronics and the continuous evolution of wind turbines, marked by improvements in aerodynamics, all supported by massive investment policies in this sector, make this sector competitive in terms of production and pricing

Wind energy production has continued to increase, reaching a record level in 2018 of 1,838.26 terawatt-hours (TWh), which is approximately 16% more than production in 2020 (Statista, 2023). If we look back to 2010 when production was 342,83 TWh, we see an increase of approximately 436,20%. According to the Global Wind Energy Council report (2024), the year 2023 was a record year with a capacity of approximately 117 GW of installed wind energy; this is 50% more than production in 2022. Wind energy production occurs in both onshore and offshore environments, with offshore installations being implemented in the last decade. Between 2010 and 2020, onshore production increased from 178 GW to 699 GW, while offshore production increased proportionally more but from a lower base (from 3,1 GW to 34,4 GW). Offshore energy, still underutilized, remains a significant potential for global production growth. At the same time, the prices of these energy sources have continued to decline over time. The main issue with this energy was its high cost, limiting its accessibility. Today, this hurdle seems to have been overcome thanks to various support policies for the sector. The installation costs for wind energy are now more or less in line with those of fossil fuels, and sometimes even lower. While the production cost was around 0,55 USD/kWh in 1980, wind projects today produce electricity at prices ranging from 0,05 to 0,09 USD/kWh without financial support (IRENA, 2021). It is reported that the production costs for both onshore and offshore have decreased from 68 to 60 between 2010 and 2021 (Osman et al., 2023).

Forecasts predict that production costs will decrease to 0,03 USD/kWh by 2030 (IEA, 2021), thanks to technological advancements and increased scale policies. Furthermore, the reduction in these production costs would lead to an increase in production, which would allow the renewable sector to provide 65% of the world's electricity, with the wind sector contributing 30% (Osman et al.,2023).

Long before photovoltaic energy and wind energy, hydropower distinguished itself as the most used energy in the category of renewable energies.

1.4.3 Hydropower

Like wind energy, hydropower is also one of the oldest renewable energy sources, but it uses water as its primary source instead of wind. Its system is based on a regular mechanism driven by water pressure, current velocity, and the fall of water on the turbine invented by L. Euler in the 18th century. Today, the new technology outlined by Hansen (2016) identifies four main elements used in hydropower plants: a dam that retains water and creates a difference to generate a waterfall; a turbine that converts kinetic energy into mechanical energy; a generator that transforms mechanical energy into electricity; and a transformer that stores the electrical voltage for transmission over relatively long distances (Brookshier, 2004; Hansen, 2016). The literature also identifies three major types of hydropower plants: run-of-the-river plants that utilize the constant flow of rivers without creating large reservoirs; reservoir plants that store water in large reservoirs and release it to produce electricity according to demand; and pumped storage plants that operate like giant batteries, pumping water to an upper reservoir during low-demand periods and releasing it to generate electricity during high-demand periods (Gulliver & Arndt, 1991). The advantage of this energy lies in its stability and continuity. Furthermore, the operating and maintenance costs are very low once installed (IEA, 2020). This reduction in costs has facilitated an increase in production over time.

Today, hydropower accounts for about 37% of the renewable energy mix, compared to 31% for solar energy and 27% for wind energy (IRENA, 2023). Its share in global electricity production is 15%, making it the third-largest source of electricity worldwide, behind coal (41%) and gas (21%). Recent figures indicate a production of approximately 4,300 TWh of electricity from this source, representing a 26,46% increase from 2009 (3,400 TWh) (IEA, 2024). However, dams remain underutilized. In South America, 30% of hydropower potential is exploited, in Asia (22%), and in Africa (7%). The goals aim for a tripling of these rates by 2100.

At the same time, the production costs of hydropower require meticulous planning to reduce construction and equipment costs, which can be high. They generally vary according to the type and size of the hydropower plant, but estimates indicate that startup costs are distributed as follows: civil engineering works (40%), turbine and generator (30%), control equipment (22%), and management costs (8%) (Osman et al., 2022). The average global cost of hydropower production increased from USD 1,315/kW in 2010 to USD 2,018/kW in 2022. Considering the latter price in 2022, Naranjo-Silva et al. (2022) even found that it is very minimalistic, as some productions could cost between 12% and 119% more than the IRENA average. Macroeconomically, the average cost of electricity from hydropower sources is estimated at USD 0,048/kWh, which is 24% more than the price in 2010 (LCOE = USD 0,039/kWh) (IRENA, 2022).

Therefore, even though the production of hydropower is steadily decreasing, we see that installation, maintenance, and production costs are continually increasing.

1.4.4 Biomass Energy

Biomass is also a promising energy source in the category of renewable energies. According to Law 5346 of May 10, 2005, in the Republic of Turkey, biomass is defined as "municipal waste (including landfill gases), as well as waste vegetable oils, agricultural waste without human or animal value, resources obtained from by-products resulting from the processing of forest products other than industrial wood, tire waste, and industrial resources, including residual sludge and treatment sludges." The production of this energy relies on the combustion of various materials such as wood, plants, agricultural waste, and biogas resulting from the fermentation of these materials. Biomass is distinguished into dry or lignocellulosic types (wood, straw, herbaceous species, energy crops, etc.) and wet types (agricultural waste and agri-food waste for 96%, sewage treatment plant waste, and household garbage for 4%) (Fontecave and Candel, 2024). Organic materials allow for the generation of three types of energy: heat, electricity, and motive power. Unlike coal, its low sulfur and nitrogen content reduces the emission of gaseous pollutants during combustion (McKendry, 2002; Balat et al., 2009; Vassilev et al., 2010). Biomass is considered carbon neutral because the CO2 released during its energy conversion through combustion corresponds to the amount of CO2 absorbed by photosynthesis during the plant's growth.

Biomass is one of the main renewable sources used to generate heat (70%) (IEA, 2015). Although its share remains low in the subsector of global electricity production (2.4% in 2022) and in the energy mix (5.02% in 2019), it is nonetheless significantly utilized for heating and cooking in households (54%) and for heat in industry and buildings (21%) (IEA, 2021a; IRENA, 2022a; Global Electricity Review, 2024). Current projections indicate that bioenergy could reach 25% of the energy mix and 17% of global energy consumption by 2050. This would represent 153 EJ, or three times more than in 2018 (54 EJ) (IRENA, 2022).

1.4.5 Geothermal Energy

Geothermal energy encompasses all technologies capable of harnessing energy or heat released from the Earth, whether from its surface or from deep underground. The Earth's core consists of magma, which constantly produces intense heat moving from the center to the surface. In the first few hundred meters of depth, there is an increase in subsurface temperature, averaging about 3°C per hundred meters. Geothermal energy is used for heating buildings, generating electricity, and supporting industrial processes (Barbier, 2002; Archer, 2020). Heat beneath the Earth's surface is captured and utilized either directly or through a heat pump (HP) if the temperature level is insufficient. Geothermal resources with temperatures exceeding 110ºC can produce heat and can also be combined with heat production. When energy is extracted from

depths between 0 and 200 meters, it is referred to as shallow geothermal energy, while beyond this depth, it is known as deep geothermal energy (Bronicki, 2013).

1.5 Fossil and renewable energy issues, policies and objectives

Energy policies have evolved over time, allowing for the establishment of short-, medium- and long-term energy objectives. However, there is one issue on which they all agree: the production and consumption of fossil fuels must be significantly reduced in favor of renewable energy sources.

Current policies, largely focused on combating climate change, are based on developing renewable energy, improving energy efficiency in technologies, and fostering a culture of energy conservation. These scenarios depend on a shift in energy governance, aiming by 2050 to reduce fossil fuel consumption by two-thirds, significantly decrease carbon emissions, and gradually phase out oil, gas, coal, and nuclear energy. These energy transition projects seek a paradigm shift towards a low-carbon society. However, achieving these goals requires collective awareness and strong political will (Zelem, 2012).

As early as 1992, 178 countries at the UN Conference ratified the Rio Declaration (Climate Convention), highlighting the need to cut greenhouse gas emissions to reduce humanity's impact on climate change (Jouzel and Euvé, 2016). The annual COP (Conference of the Parties) meetings began in 1995 in Berlin, where each country or region was given specific emission reduction targets. The 1997 COP marked a significant milestone with the introduction of the Kyoto Protocol, a framework aimed at regulating CO_2 emissions across more than a hundred countries. The targets were clear: to reduce global greenhouse gas emissions by 5,2% by 2020, using 1990 as the baseline. For the European Union, this goal was set at an 8% reduction. Although it took longer than expected, the protocol was officially ratified in 2005 at COP11 in Montreal, but without the participation of the two largest greenhouse gas emitters, China and the United States.

After the Kyoto Protocol, the most significant COP was the 2015 COP21, which resulted in the Paris Agreement. This agreement aimed to be universal, bridging the gap between developed and developing countries. It commits the international community to doing everything possible to keep global warming below 2°C, ideally limiting it to 1.5°C (Ben Ahmed et al., 2003). To achieve this goal, all countries set national targets involving community actors, businesses, and citizens. However, when adding up the national contributions, global warming is projected to rise by 3,5°C by 2100, which, according to international estimates, is far from the agreed targets. The international goals stipulate that renewable energies should account for 66% of the energy mix by 2050, with renewable electricity making up 64% of the sector (Tétart, 2023).

As part of the Climate-Energy package, the European Union, for example, adopted a directive in 2009 to promote renewable energy use. The first goal was to reduce greenhouse gas

emissions by 20% by 2020, and the second was to achieve 20% renewable energy in total energy consumption by the same year (Collard, 2015). However, these targets were updated by the European Council in 2020 with new figures aiming for a 40% reduction in greenhouse gases from 1990 levels and a 27% share of renewable energy in final consumption by 2030 in the European zone (CAREY, 2024).

Similarly, China, one of the world's largest greenhouse gas emitters, set a target in 2015 to reach a 20% share of non-fossil energy in primary energy consumption by 2030 compared to 2005 levels, and to increase forest volumes by 4,5 billion cubic meters from 2005 levels (Wang and Wang, 2015).

Although actions to achieve short- and long-term objectives are increasing, literature has highlighted several key areas that need to be addressed. First, the removal of subsidies for fossil fuels. Many countries try to keep fossil fuel prices below international levels through massive subsidy policies. This policy, therefore, hampers energy efficiency, leads to overconsumption of fossil fuels, and obstructs the penetration of renewable energies in the pursuit of zero-carbon goals (Moltke et al., 2017). In 2016, the IEA estimated these subsidies at $550 billion in 2013, which is four times higher than subsidies for renewable energy (Clements et al., 2014; Collard, 2015). This figure is even more staggering according to the IMF, which put it at $4,9 trillion (Collard, 2015). Finon (2016) emphasized that although this policy has been adopted by countries since 2009, it has struggled to gain momentum, and efforts should focus on this aspect in the fight against climate change.

Secondly, despite the considerable reduction in renewable energy costs, these technologies remain very expensive. These high prices would not even favor the acquisition of renewable energy in certain parts of the world, such as in African countries (Diop, 2015). As such, these developing nations can only rely on international community support to be involved in this climate revolution.

All things considered, international policies aim to combat climate change by promoting green energy and reducing the consumption of fossil fuels. Their challenges focus on limiting global warming to below +2°C, or even 1,5°C, by the middle of the century. By 2030, renewable energy should account for more than 30% of global primary energy consumption, and over 60% by 2050. Greenhouse gas emissions are expected to decrease by 40% by 2030, based on 1990 levels. To achieve these goals, actions have been divided according to each nation that ratified the agreements, and these actions have been used to develop national action plans for each state that has committed to the agreements.

1.6 Conclusion

The analysis of the global energy landscape reveals a profound transformation underway in the structure of energy supply. While fossil fuels continue to hold a dominant position, their future

is jeopardized by the depletion of reserves and the environmental impacts they generate. In response to this critical situation, renewable energies are experiencing remarkable growth, supported by technological advances, gradually declining production costs, and increasing political commitment at the international level. However, despite these significant advances, the share of renewable energy remains relatively limited in the global energy mix. Major challenges persist, particularly in terms of investment, technological accessibility, and equity between countries. This situation highlights the need for ambitious and well-coordinated public policies to accelerate the energy transition. Since such policies often rely on short- and medium-term objectives, they require close monitoring, which calls for accurate forecasting of both fossil and renewable energy trends. Chapter 2 focuses on the models that can be used to predict the evolution of these two types of energy

CHAPTER II: Artificial Neuronal Networks

2.1 Introduction

Artificial Neural Networks (ANNs) represent new approaches, enabled by computational methods, that allow tackling problems such as classification, categorization, pattern recognition, associative memory, and predictions using artificial cells (Drew and Monson, 2000; Crone, 2005). Belonging to the family of machine learning, ANNs are inspired by the structure and function of biological neural networks in the human brain, with interconnected neurons (see Figure 1). It is important to note that neurons consist of billions of cells, and each neuron is made up of a cell body that processes information by transmitting it to and from the brain, as illustrated in Figure 1 (Van Gerven and Bohte, 2017).

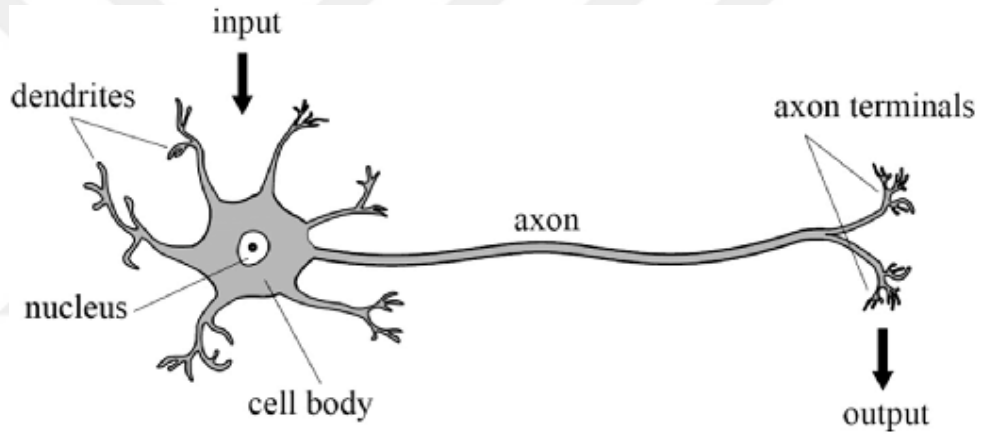


Figure 1: The biological neuron structure (Neves et al., 2018).

Thus, the human neural system is based on four fundamental elements, namely:

- **Dendrites**: These receive information or signals from other neurons.
- **Nucleus and Cell Body**: These are responsible for processing the information received from the dendrites.
- **Axon**: the element that carries information from one neuron to another.
- **Axon terminals**: specialized region at the end of an axon that contains finger-like projections. It is responsible for releasing neurotransmitters.

Artificial Neural Networks (ANNs) are generating growing interest in the field of predictive modeling, particularly due to their ability to mimic the functioning of the human neural system. Like biological neurons, ANNs can receive signals, processing them, and producing responses, thereby forming a dynamic and adaptive learning system. According to Dastres and Soori (2021), the fundamental objective of these structures is to develop automated systems capable of efficiently processing data and generating new knowledge from raw information. This process is based on the ability of ANNs to model complex relationships between variables without requiring strict prior assumptions about data distribution. In this respect, ANNs stand out as one of the most powerful approaches to information processing, especially for identifying nonlinear structures, recognizing patterns, and predicting time series.

In this chapter, we present a detailed explanation of how these ANNs function, their main variants (such as MLP, LSTM and ARIMA-ANN), and how they are applied to time series modeling and forecasting.

2.2 Why ANNs

The success of ANNs is based on the fact that they can estimate a large family of real functions that depend on explanatory or input variables, generating by machine learning target functions that can approximate output values with a desired degree of precision, unlike other classical methods. Taking classical regression methods as an example, the drawback that has long hampered researchers is the need to create assumptions about the data. These assumptions may relate to linearity, normality, homoscedasticity, and assumptions about errors that must be independent. Results without these assumptions are generally considered unreliable. Additionally, another disadvantage of linear models is multicollinearity among the predictor variables. When multiple variables in a linear regression model are correlated, it leads to unreliable estimates and sometimes bias in the predicted results (Shams et al., 2021). On the other hand, ANNs, through their algorithms, can recognize hidden patterns and correlations in raw data, group them, learn, and continuously improve until the optimal model is found. Ghorbanzadeh (2024) highlighted several advantages of ANNs over linear modeling in general:

- **Data Handling**: ANNs handle missing or noisy data better than multiple regression models.
- **High-dimensional data handling**: Multiple regression models can become cumbersome with many input factors, while ANNs can process massive datasets with large quantities of inputs and extract meaningful results.
- **Better generalization**: ANNs perform better on unseen datasets compared to multiple regression models.

- **Stability and flexibility**: The stability and flexibility of ANNs allow them to accept new information without losing what has already been learned.

2.3 Stucture of artificial neural networks

According to Brouillard (2020), a neural network is made up of several layers, each representing a mathematical operation. This is a function $f: \mathcal{X} \to \mathcal{Y}$, which is itself a composition of several functions $f^{(k)} \circ \cdots \circ f^{(2)} \circ f^{(1)}(x)$ corresponding to the different layers (see Figure 2). Each layer *i* is parameterized by a bias $\boldsymbol{b}_i \in \mathbb{R}^{h_{i+1}}$ and a weight matrix $W_i \in \mathbb{R}^{h_{i+1} x\, h_i}$ where h_{i+1} corresponds to the dimensionality of the output of $f^{(i)}$and h_1 the dimensionality of the input *x*. Each $f^{(i)}$ is a linear transformation followed by a nonlinearity φ_i that is applied to each element,

$$f^{(i)}(\boldsymbol{x}) = \varphi_i(W_i \boldsymbol{x} + \boldsymbol{b}_i) \tag{1}$$

with $1 \leq i \leq k$

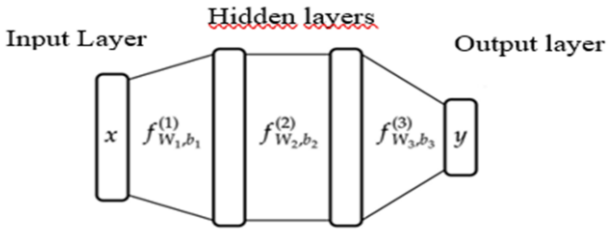


Figure 2: Representation of an artificial neural network with two hidden layers

The neural network is usually denoted by $f_\theta \, where \, \theta = \{W_i, \, b_i\}_{1 \leq i \leq k}$ is the list of all parameters for the ***k*** layers. The latter is determined by the type of problem considered. By way of illustration, for a linear regression, the last layer can only be a linear application (without the φ_i non-linearity). For a classification problem, the *softmax* σ function is used to return values similar to probabilities (i.e. in the interval [0,1] and summing to 1). The *softmax* function is defined as follows on a vector

$$\sigma(z_1, \cdots, z_d) = \frac{1}{Z}(e^{z_1}, \cdots, e^{z_d}) \tag{2}$$

where $Z = \sum_{i=1}^{d} e^{z_i}$

As mentioned above, the composition of ANNs is similar to biological neural networks. They are made up of inputs, weights, a summation function and an activation and output function (see Figure 3).

Thus, if we consider neuron j in layer C^l , neuron activation, denoted ϑ_j^l is calculated so that.

$$\vartheta_j^l = g\left(z_j^l\right) \tag{3}$$

$$z_j^l = \sum_{i=1}^{n} w_{ij}^l \vartheta_i^{l-1} + b_j^l \tag{4}$$

- z_j^l : the post-synaptic potential of neuron j in layer C^l,
- w_{ij}^l : the weight linking neuron i in layer C^{l-1} to neuron j in layer C^l,
- b_j^l : the bias of neuron j in layer C^l,
- g is the continuous activation function.

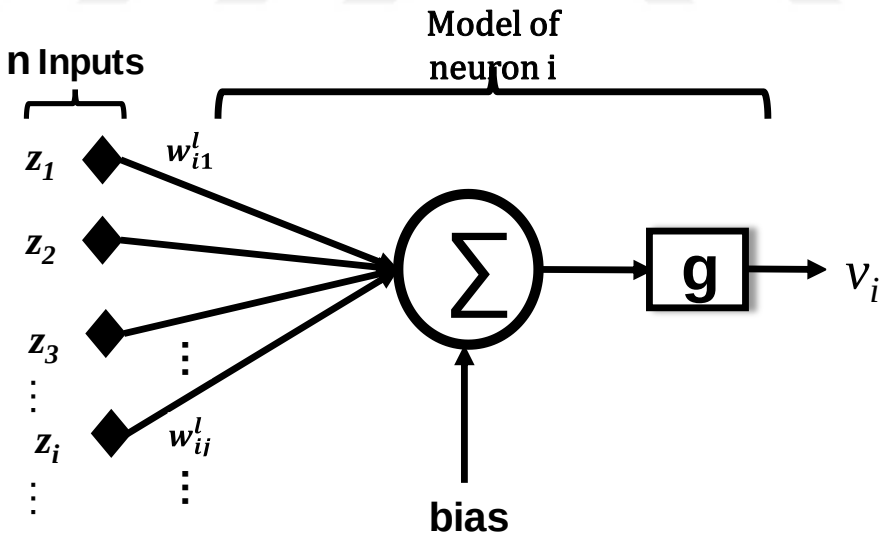


Figure 3: Structure of an ANN

The input layers can consist of one or more inputs, depending on the research objective.

The adaptive weights (w_{ij}) that interconnect the artificial neurons in the hidden layers and the output layer are calibrated by a training process with input-output data (Bre et al., 2018)

Next, the data processing process in ANNs integrates activation functions. Neural networks without activation functions would be limited to linear transformations, making them unsuitable for solving complex problems. The activation function is the element that introduces non-linearity into neural networks, captures complex patterns, and provides more accurate predictions. An activation function is associated with each artificial neuron, and it can be any function with values in the range of [-1; 1]. These functions, which have evolved over time, initially included Heaviside or threshold functions, which only took two values: 0 and 1, depending on whether the strength of the signal arriving at the neuron exceeded a certain threshold, with the threshold value being the opposite of the bias. Unfortunately, the drawback of this function is that it is not differentiable over $\mathbb{R}$. This leads to difficulties in determining the weights and biases during the learning phase. To address this non-differentiability problem, sigmoid functions, which approach the Heaviside function, are used (Voyant, 2011; Lamy, 2022). Table 1 lists some activations functions mostly used in neural networks architecture.

Table 1: Some activation functions and their derivatives.

Function	*Equation*	*Derivative*
Heaviside ou threshold	$h(z_j^l) = \begin{cases} 1, & z_j^l \geq 0 \\ 0, & z_j^l < 0 \end{cases}$	$h'^{(z_j^l)} = \partial(z_j^l)$
Sigmoïd	$\sigma(z_j^l) = \frac{1}{1 + e^{-z_j^l}}$	$\sigma'(z_j^l) = \sigma(z_j^l)(1 - \sigma(z_j^l))$
Hyperbolic tangent	$tanh(z_j^l) = \frac{e^{z_j^l} - e^{-z_j^l}}{e^{z_j^l} + e^{-z_j^l}}$	$tanh'(z_j^l) = 1 - tanh(z_j^l)^2$
Softmax	$\mathcal{S}(z_j^l) = \frac{1}{\sum_{k \epsilon \mathcal{C}^l} e^{z_k^l}}$	$\mathcal{S}(z_j^l) = \mathcal{S}(z_j^l)[1 - \mathcal{S}(z_j^l)]$
Rectified Linear Unit (ReLU)	$ReLU(z_j^l) = \begin{cases} z_j^l, & z_j^l \geq 0 \\ 0, & z_j^l < 0 \end{cases}$	$ReLU'(z_j^l) = \begin{cases} 1, & z_j^l \geq 0 \\ 0, & z_j^l < 0 \end{cases}$
Identity	$I(z_j^l) = z_j$	$I'(z_j^l) = 1$

2.4 Main architectures of ANNs

As soon as two neurons are connected, we can speak of a network. It's the way in which these neurons are organized that determines the architecture of the network. Defining the architecture of a neural network relies on a judicious choice of input vector, network size (number of layers and neurons), network structure (types of interconnection between layers of neurons) and transfer functions, without having a standard procedure (Coulibaly et al., 1999). Defining network size and structure are very difficult steps in the construction of ANNs. However, the literature shows that there are three main forms of ANNs: Feedforward Neural Networks (FNN), Recurrent Neural Networks (RNN) and Convolutional Neural Network (CNN).

2.4.1 Feedforward Neural Networks (FNN)

In this ANN, the information flows in one direction. It moves forward without feedback loops, passing through the input nodes, then the hidden nodes (if any), and ends at the output nodes. The inputs and outputs are fixed in these cases. Each layer consists of a number of neuron variables. In this type of architecture, Multilayer Perceptrons (MLPs) or Deep Neural Networks (DNN). MLPs were developed to address the problem of single-layer networks, which only have an input and output layer and can only learn linear decision boundaries. Multilayer neural networks, on the other hand, are used for more complex tasks such as classification and regression (Montufar et al., 2014; Hoang et al., 2021).

2.4.2 Récurrent Neural Networks (RNN)

Recurrent Neural Networks (RNNs), unlike FNNs, are characterized by recurrent loops in their architecture. This allows them to identify patterns or models in sequential data by taking into account contextual information provided by previous segments for prediction or classification. As a result, this structure can retain input data. Recurrent Neural Networks (RNNs) are designed to leverage sequential data by learning from previous observations to predict future patterns. Consequently, past information must be retained to make accurate forecasts. In RNNs, the hidden layers serve as internal memory, preserving key details from earlier stages of processing sequential inputs. This type of network is typically used for textual data, speech, or time series. Virtual applications like Siri and Cortana illustrate practical applications of this network architecture. However, research has shown that RNNs are less commonly used for regression tasks. The main obstacle of RNNs is the vanishing gradient phenomenon, where the gradient of the loss function becomes too small. This leads to a problem with learning long-term dependencies. The complexity of learning in RNNs is another challenge, especially when the dataset is relatively large, making this architecture less practical for certain types of prediction (Ghorbanzadeh, 2024).

2.4.3 Convolutionnal Neural Networks (CNNs)

Convolutional neural networks (CNNs) use convolution as a basic function instead of matrix-vector multiplication as in MLPs. To achieve invariance, they use parameter sharing and pooling operations. They are commonly used in computer vision tasks, object detection, and image classification tasks. Additionally, radial basis function (RBF) networks rely on functions that detect points affected by noise. The universal approximator for this type of network uses Gaussian activation functions (Voyant, 2011). However, since this architecture is not closely related to our research object, we will not explore it in depth.

2.5 Neural Network Approaches to Time Series Forecasting

Before diving into the approach of modeling temporal data, it is important to highlight an empirical approach to modeling times series data.

2.5.1 Auto-Regressive Integrated Moving Average (ARIMA)

The integrated Autoregressive Moving Average (ARIMA) is an algorithm that has demonstrated its performance in terms of accuracy in forecasting time series (Siami-Namini, 2018). This classical statistical model is a combination of other traditional methods such as auto-regressive (AR) and moving-average (MA) processes. Better generalized by the Box-Jenkins approach (Anderson, 1977), this stochastic process is particularly useful for forecasting linear time series data. Considering a time series x_t, we have:

- *AR* (*Autoregression*): Captures the dependency between an observation and its past values (lags). Let p be the number of observations the process takes. We have AR, denoted AR(p) model of order p be:

$$x_t = c + \sum_{i=1}^{p} \emptyset_i x_{t-i} + \epsilon_t \tag{5}$$

- MA *(Moving average)*: An approach that accounts for the relationship between observations and residual error terms when using a moving average model on lagged observations (q). The MA of order q ; i.e. MA(q) can be written as follows:

$$x_t = \mu + \sum_{i=0}^{q} \theta_i \epsilon_{t-i} \tag{6}$$

μ is the expectation of x_t assuming equal to zero.

Hence, the ARIMA model of order (p, q) is then:

$$x_t = c + \sum_{i=1}^{p} \emptyset_i x_{t-i} + \epsilon_t + \sum_{i=0}^{q} \theta_i \epsilon_{t-i} \tag{7}$$

where $\emptyset_i \neq 0, \theta_i \neq 0$ and $\sigma_\epsilon^2 > 0$

- The parameters p and q are called the AR and MA orders
- The θ_i terms are the weights applied to the current and prior values of a stochastic term in the time series, and $\theta_0 = 1$.
- $\emptyset_i$ with $1 \leq i \leq p$ are autocorrelation coefficients
- We assume that ϵ_t is a Gaussian white noise series with mean zero and variance σ_ϵ^2
- c is a constant

The general form of a seasonal ARIMA model is denoted as ARIMA (p, d, q) where ***d*** is the non-seasonal differencing. The most important step in estimating seasonal ARIMA model is to identify the values of (p, d, q) (Siami-Namini, 2018)**.**

By observing time graphs and using the autocorrelation function (ACF), it is possible to visualize the trend and seasonality of the data, and to measure the degree of linear dependence between observations in a time series that are separated by a lag p. This will enable variance-stabilizing transformations and differentiations to be used, and the value of p to be estimated. The partial autocorrelation function (PACF) is used to determine the number of autoregressive terms q. The order of differentiation of d is generally the frequency difference between a non-stationary and a stationary time series. This is illustrated by Equation (8):

$$\nabla^d x_t = \nabla^{d-1} x_t - \nabla^{d-1} x_{t-1} \tag{8}$$

where ∇^d is the d-order difference.

It should be noted that a high number of differentiations results in a large loss of data. This is why, for each differentiation, it is necessary to carry out a statistical test such as an ADF (Augmented Dickey-Fuller) unit root stationarity test. The difference at which the test result is satisfactory expresses the value of parameter *d.*

2.5.2 ANN approach to time series modeling

ANNs are computational methods that can be used in a wide range of problems, particularly those involving time-series forecasting. ANN models have the advantage of approximating a large class of nonlinear functions with methods that do not require a priori assumptions compared with other nonlinear models (Zhang and Min Qi, 2005). This method therefore compensates for the shortcomings of conventional stochastic methods, which usually only interpret linear data.

There are several approaches to finding optimum architecture, and the literature has enabled us to summarize them as follows[1]:

- Methods based on direct transmission of information from input to output. These methods use architectures such as Multi-layer Perceptrons (MLP), Deep Neural Network (DNN) and Radial Basis Function Network (RBFN), etc.
- Methods that call on architectures with recurrent loops and the ability to retain information passed to their hidden neurons. LSTM networks, Vanilla RNN, Gated Recurrent Network (GRU), etc.
- Hybrid methods where ANNs are combined with other techniques such as fuzzy systems, evolutionary algorithms or probabilistic models. These systems aim to leverage the advantages of each method to create more efficient solutions to complex problems. This method will be described in more detail below taking an exemple (ARIMA-ANN, ConvLSTM, etc).

Since this study focuses mainly on time-series data with matrix inputs and output, we will focus on describing Multi-layer Perceptron (MLP), hybrid systems (ARIMA-ANN) and LSTM architecture.

2.5.2.1 Multi-layer Perceptron (MLP)

Multi-layer feedforward networks are the most widely used networks in time-series forecasting (Jain et al. 1996; Crone, 2005; Hwang et al. 2006). The relationship between outputs x_t and inputs $x_{t-1,\cdots,}x_{t-p}$ is expressed by the following mathematical function:

[1] *For a more detailed summary, please refer to articles by Lim and Zohren (2018), Torres et al. (2021) who provide a brief overview of these different groups.*

$$x_t = w_0 + \sum_{j=1}^{q} w_j \,.\, g\left(w_{0,j} + \sum_{i=1}^{p} w_j \,.\, w_{i,j} \,.\, x_{t-i}\right) + \varepsilon_t \tag{9}$$

where, $w_{i,j}(i = 0,1,2,\dots,p, j = 1,2,\dots,q)$ and $w_j(j = 1,2,\dots,q)$ are model parameters often called connection weights; p is the number of input nodes; and q is the number of hidden nodes (Zhang, 2003). There is no standard method for determining the q value, which depends fundamentally on the size of the data. Furthermore, one of the parameters to be considered in a network is the number of shifted observations, p the dimension of the input vector. Zhang (2003) considers this to be the most important parameter in an MLP model, as it plays a crucial role in determining the (non-linear) autocorrelation structure of the time series. Once the network structure (p, q) has been defined, training is carried out to estimate the parameter values.

In general, the standard form of the ANNs model resulting from Equation (10) implies a nonlinear functional mapping from past observations$\left(x_{t-1}, x_{t-2}, x_{t-p}\right)$ to the future value x_t:

$$x_t = f\left(x_{t-1}, x_{t-2}, x_{t-p}, w\right) + \varepsilon_t \tag{10}$$

where

- f is an activation function determined by the network structure and connection weights.
- w is a vector of all parameters

2.5.2.2 Long Short-Term Memory (LSTM)

2.5.2.2.1 LSTM Classic

The LSTM is a form of RNN with additional features to memorize the data sequence and able to learn long-term dependencies. It was developed to overcome the inherent limitations of classical RNNs, particularly in terms of long-term memory management (Hochreiter and Schmidhuber, 1997).

What fundamentally sets the LSTM model apart is its ability to retain, update, and selectively retrieve information across long sequences, making it a preferred choice for processing sequential data. This makes it particularly relevant for tasks involving temporal or ordered data.

It relies on the integration of control mechanisms, known as gates, which finely regulate the transfer, forgetting and retention of information over time. All gates are linked to a memory cell

capable of capturing and storing information received from the gates. The cell state vector acts as the primary channel through which information flows, allowing it to pass forward with minimal alterations. However, linear transformations can be applied when necessary. This flow can be modulated either enhanced or diminished through the action of sigmoid gates. These gates, akin to neural layers or successive matrix operations, are equipped with distinct weights that enable precise control over the information being processed (Yu et al., 2019).

There are 3 main gates in a block: a forget gate, an input gate and an output gate as shown in Figure 4.

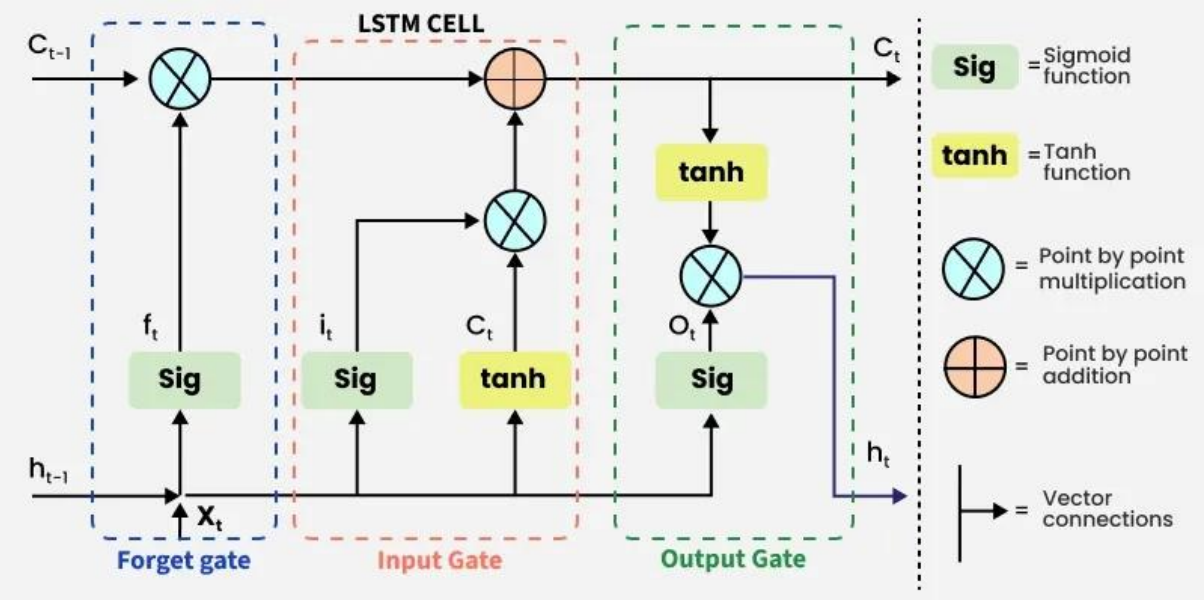


Figure 4: LSTM Bloc.

- **Forget gate** (G_t) : selects which information from the unit's previous state should be retained or discarded in the current state. This selection and deletion mechanism is controlled by the sigmoid function, which takes as input the previous hidden state of the LSTM unit $[h_{t-1}]$ and the current input X_t. It determines which components of the previous cell state should be retained or discarded. The forget gate, denoted as G_t, is a vector with values ranging from 0 to 1, each corresponding to a component of the cell state C_{t-1} (Kumar et al., 2023). The mathematical equation for this gate is as follows:

$$G_t = \sigma(W_G . [h_{t-1}, x_t] + b_G) \quad (11)$$

Here, σ is the sigmoid function, and W_G and b_G are the weight matrices and bias, respectively of the forget gate.

- **The input gate** (i_t) determines which part of the new X_t input information is to be incorporated into the current memory. This step is subdivided into two layers: It also relies on a sigmoid function (0 or 1), which selects relevant components from the current input X_t and the previous hidden state h_{t-1}. These components are then passed through a *tanh* function (-1 to 1) to produce a vector of candidate values $(\widetilde{C_t})$, which are partially integrated into the cell state based on the weights computed by the input gate, deciding on their level of importance (Kumar et al., 2023). This new memory is then added to the old memory C_{t-1} resulting in C_t.

$$i_t = \sigma(W_i \,.[h_{t-1}, x_t] + b_i) \tag{12}$$

$$\tilde{C}_t = tanh(W_C \,.[h_{t-1}, x_t] + b_C) \tag{13}$$

$$C_t = G_t \,.\, C_{t-1} + i_t \,.\tilde{C}_t \tag{14}$$

where C_{t-1} and C_t are the cell states at time $t-1$ and t, while W_i, W_C and b_i , $b_c$are the weight and bias matrices, respectively, of the cell state.

- **The output gate** (O_t) : decides to what extent the contents of the cell influence the output produced by the unit. The output gate controls the information extracted from the cell state to generate the LSTM unit's output at time step *t*. A sigmoid function is applied to h_{t-1}. and X_t to determine which parts of the cell state (C_t) contribute to the output. The updated cell state is passed through a *tanh* function, and the result is then element-wise multiplied by the output gate vector, yielding the final output h_t.

$$O_t = \sigma(W_o \,.[h_{t-1}, x_t] + b_o) \tag{15}$$

$$h_t = O_t \,.\tanh(C_t) \tag{16}$$

Here, W_o and b_o are the weight and bias matrices, respectively, of the output gate.
Each LSTM unit has a memory cell with state C_t at time t, which is controlled by the above three gates.

2.5.2.2.2 Attention mechanism

The attention mechanism has emerged as one of the methods that has revolutionized neural modeling. It was recently implemented by Bahdanau et al. (2014) for machine translation using RNN and its variants but has subsequently been tested in several prediction domains (Niu et al., 2021). It is inspired by human ability, which gives more attention to important details while neglecting less important ones. They are incorporated into architecture to help artificial neural networks learn the most important information to consider when making predictions, thus reducing model bottlenecks.

Implementing the mechanism requires several orderly steps. It begins with the computation of attention scores, then the normalization of the attention score by the softmax function and finally the computation of the attention vector (Kumar et al., 2023).

(1) The attention score: Each piece of information h_t received is projected onto a scalar score α_t, using a dense layer with activation function *tanh*:

$$\alpha_t = \tanh(\mathrm{v}^T h_t + b) \tag{17}$$

where v^T and $b \in \mathbb{R}$ are weight vector and a bias that are learned. They can be jointly trained with other components of the model.

(2) Attention score normalization: The α_t scores are then normalized using the softmax function to obtain attention weights B, which is also the conditional probability distribution. So, its equation is given by:

$$\beta_t = softmax(\alpha_t) = \frac{\exp(\alpha_t)}{\sum_{t=1}^{T} \exp(\alpha_t)} \tag{18}$$

where T is the last year of the time sequence.

(3) The context vector: The attention mechanism computes the context vector z_t as a weighted sum of all hidden state h_t.

$$z_t = \sum_{t=1}^{T} \beta_t . h_t \tag{19}$$

The process of setting up the mechanism is illustrated in Figure 5.

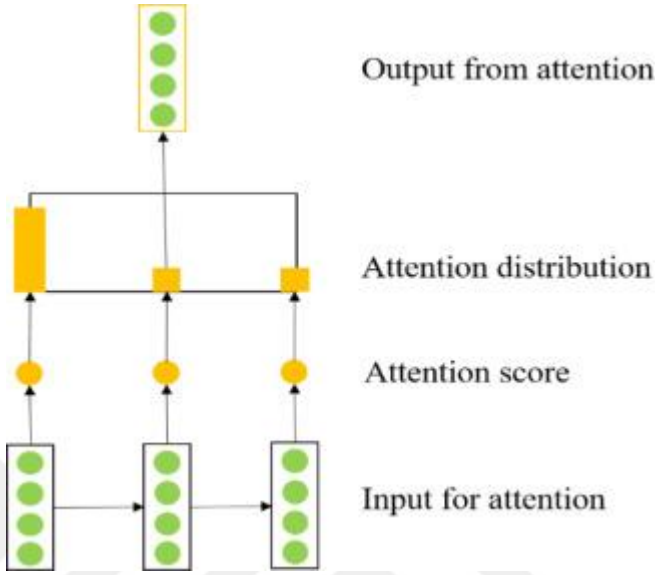


Figure 5: Attention Mechanism steps

2.5.2.2.3 Attention-based LSTM (A-LSTM) Model

Although LSTM networks have mitigated several limitations of traditional recurrent neural networks, their ability to capture long-term dependencies remains partially constrained (Pascanu et al., 2013, as cited in Kamur et al., 2023). Moreover, their predictive performance can be limited in complex scenarios where the relevance of past information varies significantly over time.

To address these shortcomings, an Attention-based LSTM (A-LSTM) model was developed by incorporating an attention mechanism into the LSTM framework. This mechanism allows the network to dynamically assign weights to each time step based on its relevance to the target prediction, thereby enhancing the model's capacity to focus on the most informative temporal features.

The A-LSTM architecture is generally composed of five sequential layers that work together to enhance the model's predictive capabilities (Kumar et al., 2023). It begins with an input layer that receives the historical observations of the time series. These inputs are passed to a standard LSTM layer, which encodes the temporal dependencies and transforms the raw sequence into high-level hidden representations. An attention layer then processes these hidden states by assigning a dynamic

weight to each time step, enabling the model to emphasize the most relevant information across the sequence. This weighted representation is then passed through a dense layer, also known as a fully connected layer, which integrates the extracted features and prepares the data for output. Finally, the output layer produces the prediction, completing the forward pass of the model. This architecture allows for more context-aware learning and enhances the model's sensitivity to temporal salience in the input data.

2.5.2.2.4 Structure of the A-LSTM model

Referring to the studies of Zhang et al. (2019) and Kumar et al. (2023) we have established the structure of the A-LSTM model.

Input Layer

Let's denote $X = (X^1, X^2, \dots, X^n)^T = (X_1, X_2, \dots, X_m) \in \mathbb{R}^{(n\,x\,m)}$ the raw times series where n is the number of feature sequences, m is the length of the window or the timestep size. In the training process of an LSTM network, or any model of its type, input data must be split into sequences. $X^k = (x_1^k, x_2^k, \dots, x_m^k) \in \mathbb{R}^m$ is used to represent a sequence (vector) of length m. We also employ $X_t = (x_t^1, x_t^2, \dots, x_t^n) \in \mathbb{R}^n$ to represent a set vector of n features at time t.

LSTM Layer

So given a sequence of inputs $(X_1, X_2, \dots, X_m)$, the LSTM compute sequence $(h_1, h_2, \dots, h_m)$, and the C-sequence $(C_1, C_2, \dots, C_m)$ as follows :

$$G_t = \sigma(W_G \,.\, [h_{t-1}, X_t] + b_G) \tag{20}$$

$$i_t = \sigma(W_i \,.\, [h_{t-1}, X_t] + b_i) \tag{21}$$

$$\tilde{C}_t = tanh(W_C \,.\, [h_{t-1}, X_t] + b_C) \tag{22}$$

$$C_t = G_t \,.\, C_{t-1} + i_t \,.\, \tilde{C}_t \tag{23}$$

$$O_t = \sigma(W_o \,.\, [h_{t-1}, X_t] + b_o) \tag{24}$$

$$h_t = O_t \,.\, \tanh(C_t) \tag{25}$$

where σ is logistic function, C_t is the cell that needs to be updated. W_G, W_i, W_C, W_o and b_G, b_i, b_C, b_o are parameters of the model and can be learned. In addition, G_t, i_t, O_t are our forget, input and output gate respectively. Each LSTM unit has a memory cell with C_t at time t, which is controlled by the above three gates.

Attention Layer

The attention mechanism enables the model to "look" at all the input data and weight their importance, in order to produce a more accurate and contextually relevant output.

Attention examines the entire sequence history and assigns an important weight to each element based on its relevance. At each prediction step, the model compares the current output (the "query") with each element of the previous sequence (the "keys"). It computes similarity scores, the more relevant an element is, the higher its score. These scores are then normalized into probabilities (ranging from 0 to 1). Finally, the model generates a context vector: a weighted sum of all past data, where each element's contribution is scaled by its assigned importance.

In the second part, we construct an attention mechanism through a deterministic attention model. For a certain feature sequence $X^k = (x_1^k, x_2^k, \dots, x_m^k) \in \mathbb{R}^m$, by referring to output h_k generated in the last LSTM unit, we define:

$$\alpha_t^k = \tanh(\mathrm{v}^T h_k + b) \tag{26}$$

$$\beta_t^k = softmax(\alpha_t^k) = \frac{\exp(\alpha_t^k)}{\sum_{k=1}^{n} \exp(\alpha_t^k)} \tag{27}$$

where the vector v and matrices W_1, W_2 are learnable parameters of the model. The vector $\alpha^k = (\alpha_1^k, \alpha_2^k, \dots, \alpha_m^k)$ has length m and its i-th item measured the importance of the k-th input feature sequence at time t. These items are then normalized by *softmax* function. β^k is an attention weight, which contains a **score** of how much attention should be put on the k-th feature sequences. We can also get the output of the attention model at time t, i.e., the weighted input feature sequence z_t as follows:

$$z_t = \sum_{k=1}^{n} \beta_t^k . h_k \tag{28}$$

Where z_t is the attention vector or context vector.

There we have the X_t in the LSTM equations replaced by newly computed z_t, to update the attention model and we finish the work of converting raw series to attention-based time series.

Traditional time series forecasting frameworks based on recurrent neural networks or LSTM architectures typically use raw time series sequences, treating all input features with equal importance. However, by introducing a context vector z_t derived from an attention mechanism, the model can selectively emphasize more relevant input features. This focused representation enhances the extraction of key patterns while eliminating the influence of redundant or less informative sequences. In theory, using $\boldsymbol{z_t}$ as the input to the deep learning LSTM model leads to improved predictive accuracy.

Dense Layer

The dense layer, commonly referred to as a fully connected layer ensures that each neuron in the previous layer is linked to every neuron in the next. Thanks to adjustable weights, it combines the dimensions of the context vector $\boldsymbol{z_t}$ to produce a final output.

Output layer

The output layer is the final layer of a neural network. Its target is to produce a prediction that matches the desired output variable (also called the label or target) of the task at hand.

The A-LSTM modeling procedure is summarized in Figure 6.

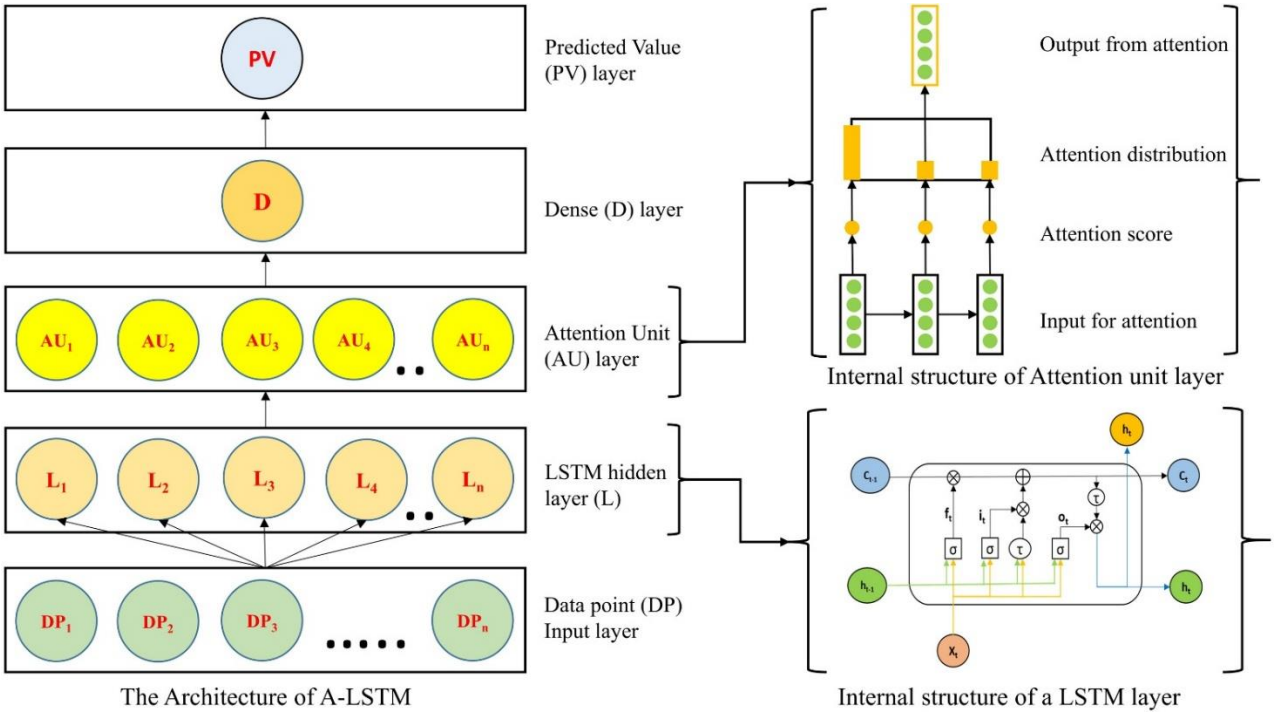


Figure 6: An illustration of A-LSTM Bloc (Kumar et al., 2023)

2.5.2.3 ARIMA-ANN model

The hybrid ARIMA-ANN model, as its name suggests, is a combination of the classical ARIMA model and artificial neural networks. The model proposed by Zhang (2003) assumes that time series data are a sum of linear and non-linear components as follows:

$$x_t = L_t + N_t \tag{29}$$

where L_t denotes the linear and $N_t$denotes the nonlinear component. However, Khashei and Bijari (2010) have questioned this additive relationship between linear and non-linear components and stipulate that it is preferable to construct a functional relationship between the components as indicated by Equation (30):

$$x_t = f(L_t, N_t) \tag{30}$$

Modeling with their method consists of 2 steps. First, use the ARIMA method to extract the linear component and make forecasts on it. Secondly, the residuals, which represent the non-linear component, are integrated into the ANN to the original data and the linear forecast from the ARIMA model. Let f be a non-linear function determined by a neural network. We have:

$$x_t = f[(z_{t-1}, z_{t-2}, \dots, z_{t-m})(e_{t-1}, e_{t-2}, \dots, e_{t-n})] \tag{31}$$

where $z_t = (1-B)^d(x_t - \mu)$, e_{t-1} is the residual at time t and m and n are integers.

Thus, the first step consists in using the ARIMA model to generate the residuals and the second step uses the ANNs to model the residuals and the original data as follows:

$$x_t = w_0 + \sum_{j=1}^{Q} w_j \,.\, g(w_{0,j} + \sum_{i=1}^{p} w_{i,j} \,.\, z_{t-i} + \sum_{i=p+1}^{p+q} w_{i,j} \,.\, e_{t+p-i}) + \varepsilon_t \tag{32}$$

where, $w_{i,j}(i = 0,1,2,\dots,p+q, j = 1,2,\dots,Q)$ and $w_j(j = 1,2,\dots,Q)$ are connection weights; p, q,Q are integers, which are determined in design process of final neural network.

2.5.2.4 The particularity of the A-LSTM model and what makes it an efficient model

Several applications have been made with this model in several fields. In finance, Zang et al. (2019) performed financial prediction of companies using data on 3 financial indices (Russell 2000, DJIA, Nasdaq). They found that the prediction of the 3 indices performed better with the A-LSTM model than with the classic LSTM and ARIMA models. Benchaji et al. (2021) improved the credit card fraud detection system with their new A-LSTM, which performed much better than the LSTM. In the energy field, Lin et al. (2022) in their zonal electric load forecast found a better with fine prediction quality using the A-LSTM model. Kumar et al. (2023) forecast oil production using well data. The model considered was the A-LSTM, which they compared with other machine

learning models. The A-LSTM model came out much better. Tai et al. (2023) also in the same vein carried out a prediction of photovoltaic energy production using several factors with the A-LSTM model. The model performed better than the LSTM, CNN or BP (BackPropagation Neural Network) models. In the field of cybersecurity, Liu et al. (2021) tested the A-LSTM model in the detection of malicious URLs. The model had an accuracy of 98.9%, which enabled them to adopt this model as capable of ensuring the security of cyberspaces. In the healthcare field, Gao et al. (2021) in predicting patients' life expectancy, and Zang (2019) in modeling patients' disease progression for prediction, experimented with A-LSTM, which gave them much better results. All this scientific research shows how dynamic the A-LSTM model can be in the modeling process.

A-LSTM networks can capture irregular trends, cycles and long temporal dependencies in time series, which can be very useful for identifying non-linear relationships between past and future data. They can also alleviate the fading gradient problem. Unlike conventional RNNs, which suffer from the problem of old information being forgotten due to the vanishing gradient, A-LSTMs can sort and retain the most relevant information making them extremely effective for predicting time series data.

Given its strengths and performance in predicting time series data, we have opted for an A-LSTM architecture for this study. The empirical construction of the model will be detailed in Chapter 3.

2.6 Conclusion

This chapter has presented Artificial Neural Networks (ANNs), highlighting their structure, operational principles, and their growing role in the field of predictive modeling. Inspired by the human brain, these networks have demonstrated a remarkable ability to process complex data, identify nonlinear relationships, and detect patterns that traditional statistical models often struggle to capture. Their ability to function without requiring strict assumptions about data distribution is a major advantage.

The chapter also introduced various types of ANN, notably MLP, LSTM, and ARIMA-ANN, in the context of time series modeling. Furthermore, one of the key innovations derived from LSTM architecture, the Attention-based LSTM (A-LSTM) was discussed to illustrate its robustness in time series forecasting. This variant will be implemented in Chapter 3 as part of the modeling of our structured energy consumption data.

CHAPTER III: Model Design and Tools

3.1 Introduction

This research work involved proposing a model for predicting global consumption of renewable and fossil energies separately. To answer the central question of whether renewable energies are on track to replace fossil fuels, we opted for an approach based on an A-LSTM (LSTM with attention mechanism) model.

The methodology adopted is based on the following stages:

- Constitution and preparation of energy data;
- Application of pre-processing techniques to ensure time series quality and stability (stationarity, normalization, etc.);
- Implementation of an A-LSTM model trained on the historical series;
- Evaluation of model performance using four metrics: MSE, MAE, RMSE and MAPE.

The aim is not only to predict future consumption levels, but also to determine whether current trends in renewable energies are sufficient to ensure their progressive substitution for fossil fuels in the short and medium term.

3.2 General architecture of the proposed model

The design of the A-LSTM model was based on several steps necessary to perform energy predictions.

First, we pre-processed the data using several methods to: (1) present and process missing values, (2) analyze the state and evolution of our raw series, (3) normalize the data, and (4) restructure the data so that they are compatible with the requirements of the A-LSTM model. These data are then integrated into the A-LSTM model for training. After learning, the generated model makes the prediction. Figure 7 shows the proposed model architecture.

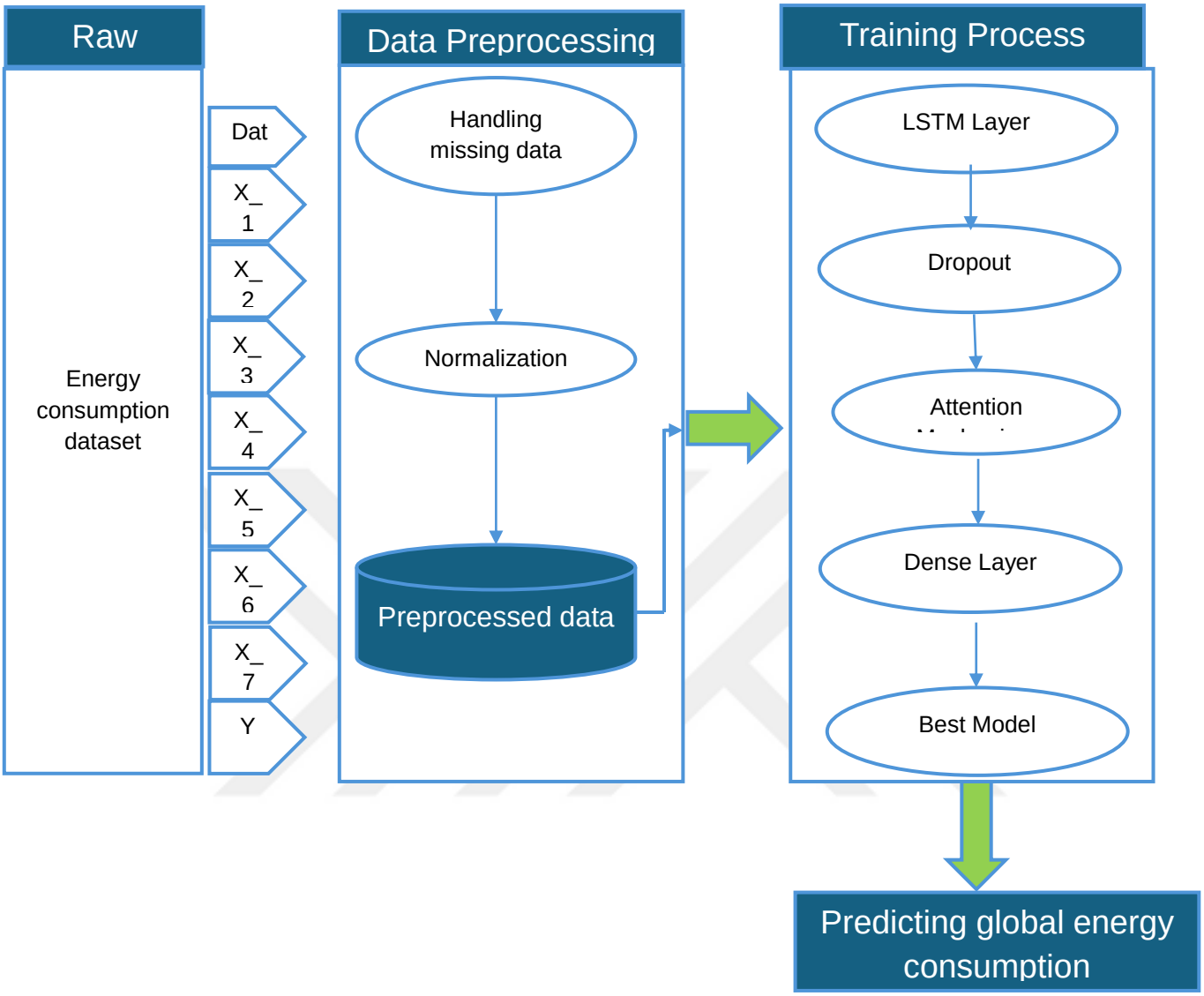


Figure 7: Architecture of our proposed model A-LSTM

3.2.1 Data Preprocessing

3.2.1.1 Description of the database

The data modelled in this research represents primary energy consumption data covering the period 1965-2022 (58 observations per series). They are subdivided into two dataset, one for renewable energies and the other for fossil energies.

Renewable energies include hydro, wind, solar, geothermal, biomass in electricity and other renewable sources. Renewable energy consumption is based on gross production and does not take into account cross-border electricity supplies. It represents the amount of fuel required by thermal power plants to produce the electricity declared.

Fossil fuels, on the other hand, take into account the consumption of oil, gas, coal and their derivatives. However, biodiesel and biogasolines are excluded, as they are now considered renewable energies. Fossil fuel consumption represents domestic demand plus international aviation and marine bunkers, fuels and refinery losses.

All the data we have used comes from the Energy Institute (EI), one of the organizations responsible for compiling the Statistical Energy Balance for the world. This institution works closely with the world's governments and energy institutions, and has been providing timely, comprehensive and objective data to the energy community since 1952. The EI Statistical Review of World Energy™ analyses data on world energy markets from the prior year. Our data comes from their Statistical Review of World Energy established in 2024, which itself collected this data from government sources and official published data[2].

The geographical area covered by the study is all the continents of the world, which we have divided into 7 regions: North America, South America, Europe, the Commonwealth of Independent States (CIS), the Middle East, Africa and Asia-Pacific. The choice of this breakdown is based on several factors. Firstly, we have the homogeneity of the energy profiles (comparable levels of development, similar energy structures, convergent energy policies) of the different countries that make up each region. For example, the CIS has a very different profile from Western Europe. Secondly, there is statistical consistency and readability between the countries in these regions, as they share similar economic and geopolitical situations. This makes it possible to standardize analyses over time, with aggregates that remain stable. Finally, the countries in each region face similar energy challenges. For example, while the Middle East is dominated by oil/gas production, Europe is moving towards a transition to renewables.

We have considered the countries covered by the EI Statistical Review of World Energy in its latest report of 2024, since our data come from this review. Table 2 lists the countries that were considered for each region during the study.

[2] *For more details, EI website is followed : https://www.energyinst.org/statistical-review/home*

Table 2: List of countries considered

North America	South America	Europe	CIS	Middle East	Africa	Asia Pacific
Canada	Argentina	Austria	Azerbaijan	Iran	Algeria	Australia
Mexico	Brazil	Belgium	Belarus	Iraq	Egypt	Bangladesh
US	Chile	Bulgaria	Kazakhstan	Israel	Morocco	China
	Colombia	Croatia	Russian Federation	Kuwait	South Africa	India
	Ecuador	Cyprus	Turkmenistan	Oman	Eastern Africa	Indonesia
	Peru	Czech Republic	USSR	Qatar	Middle Africa	Japan
	Trinidad & Tobago	Denmark	Uzbekistan	Saudi Arabia	Western Africa	Malaysia
	Venezuela	Estonia	Other CIS	United Arab Emirates	Other Northern Africa	New Zealand
	Central America	Finland		Other Middle East	Other Southern Africa	Pakistan
	Other Caribbean	France				Philippines
	Other South America	Germany				South Korea
		Greece				Sri Lanka
		Hungary				Taiwan
		Iceland				Thailand
		Ireland				Vietnam
		Italy				Other Asia Pacific
		Latvia				
		Lithuania				
		Luxembourg				
		Netherlands				
		Norway				
		Poland				
		Portugal				
		Romania				
		Slovakia				
		Slovenia				
		Spain				
		Sweden				
		Switzerland				
		Turkey				
		Ukraine				
		United Kingdom				
		Other Europe				

The two databases were built up from the data set up for each country identified by the EI structure in its latest report. The observations of a series or region in each year t for a given energy consumption represents the sum of the energy consumptions of all the countries making up this region in the said year *t*.

Let's define our raw series $X = (X^1, X^2, \dots, X^n)^T = (X_1, X_2, \dots, X_m) \in \mathbb{R}^{(n\,x\,m)}$ as the set of series for the 7 regions over the period in question (1965-2022) where $X_t = (x_t^1, x_t^2, \dots, x_t^n) \in \mathbb{R}^{1\,x\,n}$ a line vector representing the values of all regions in year t, we define :

$$x_t^k = \sum_{j=1}^{N} x_t^{k,j} \tag{33}$$

where $t = 1, \dots, m \; and \; k = 1, \dots, n$

- $n = 7$ denotes the number of variables (or regions),
- $m = 58$ denotes the number of years
- N denotes the number of countries considered in in region k
- $x_t^{k,j}$ is the energy consumption of country j in region k at year t.
- x_t^k is the energy consumption of region k at year t.
- $X^k \in \mathbb{R}^{m\,x\,1}$ is represents the full time series of region k over the observation period.

Each dataset consists of a series of eight variables, along with the year, which serves as the chronological index. Seven variables correspond to the different global regions and are used as input features, while the final variable represents global energy consumption, which serves as the output variable.

Table 3 resumes informations of our datasets.

Table 3: A summary of each dataset used in the modeling process

Data	Series	Type	Unit of measure
Renewable Data	*Year*	*Quantitative (index)*	*Date*
	North_America_Renew	*Quantitative*	*Exajoule*
	South_America_Renew	*Quantitative*	*Exajoule*
	Europe_Renew	*Quantitative*	*Exajoule*
	CIS_Renew	*Quantitative*	*Exajoule*
	Middle_East_Renew	*Quantitative*	*Exajoule*
	Africa_Renew	*Quantitative*	*Exajoule*
	Asia_Pacific_Renew	*Quantitative*	*Exajoule*
	World_Renew	*Quantitative*	*Exajoule*
Fossile Data	*Year*	*Quantitative (index)*	*Date*
	North_America_ Fossile	*Quantitative*	*Exajoule*
	South_America_ Fossile	*Quantitative*	*Exajoule*
	Europe_ Fossile	*Quantitative*	*Exajoule*
	CIS_ Fossile	*Quantitative*	*Exajoule*
	Middle_East_ Fossile	*Quantitative*	*Exajoule*
	Africa_ Fossile	*Quantitative*	*Exajoule*
	Asia_Pacific_ Fossile	*Quantitative*	*Exajoule*
	World_Fossile	*Quantitative*	*Exajoule*

3.2.1.2 Handling of Missing Values

Given that the observations in both constructed datasets are derived from country-level data aggregated by region, all preprocessing was initially carried out at the country level before regional aggregation.

In this context, we first identified the countries with missing values. These gaps were generally observed during the early years of the reference period (1965-2022). Subsequently, a set of imputation methods was applied to fill in the missing values and ensure data consistency across all regions.

- For countries with a missing data rate of no more than 5%, linear extrapolation was applied after examining the shape of their early time series, which generally exhibited linear trends. Since the extrapolation was performed backwards in time, the following formula was used:

Considering our observation $x_t^{k,j}$, the energy consumption of country j in region k at year t,

$$x_t^{k,j} = x_{t_{base}}^{k,j} - a.(t_{base} - t) \tag{34}$$

where:

- t_{base} is the first valid year and t the year whose $x_t^{k,j}$ value we have extrapolated. Ex: if values of years 1965,1966 and 1967 are missing and we have value of year 1968, 1968 will be our t_{base}
- $x_{t_{base}}^{k,j}$ is the value of t_{base}
- a a is the estimated slope between first valid year and the following year.

$a = \frac{x_{t_{base}+1}^{k,j} - x_{ij,t_{base}}}{(t_{base}+1)-(t_{base})}$

- For countries with more than 5% missing values in their time series, we opted for their exclusion to avoid introducing bias into the input data for the A-LSTM model and, more broadly, compromising the objectivity of the prediction. Countries such as China Hong Kong SAR and Singapore in the Asia-Pacific region, as well as Slovenia and North Macedonia in Europe, were therefore not included in this study.

Missing data were observed primarily in the renewable energy series. In contrast, the fossil energy data rarely contained missing values.

Furthermore, we chose to retain all countries in the CIS region (Commonwealth of Independent States). This region includes only the countries from the former USSR, excluding Ukraine, Lithuania, Estonia, and Latvia. Data for the USSR were available up to 1984, after which the successor states began publishing their own statistics. We decided to include the USSR data because our analysis focuses on the aggregate total for the CIS region rather than individual countries.

Once the missing values had been filled and certain countries excluded from the dataset, we proceeded to aggregate the totals for each region to construct the final datasets. It was then essential to analyze the overall structure and behavior of the resulting time series.

3.2.1.3 Analysis and Processing of the Different Series

As previously mentioned, the datasets consist of eight variables plus the year, including seven input variables and one output variable. These variables correspond to the world's major regions: North America, South America, Europe, the CIS, the Middle East, Africa, and Asia-Pacific. Figure 8 illustrates the evolution of renewable energy consumption in each region over the period 1965–2022.

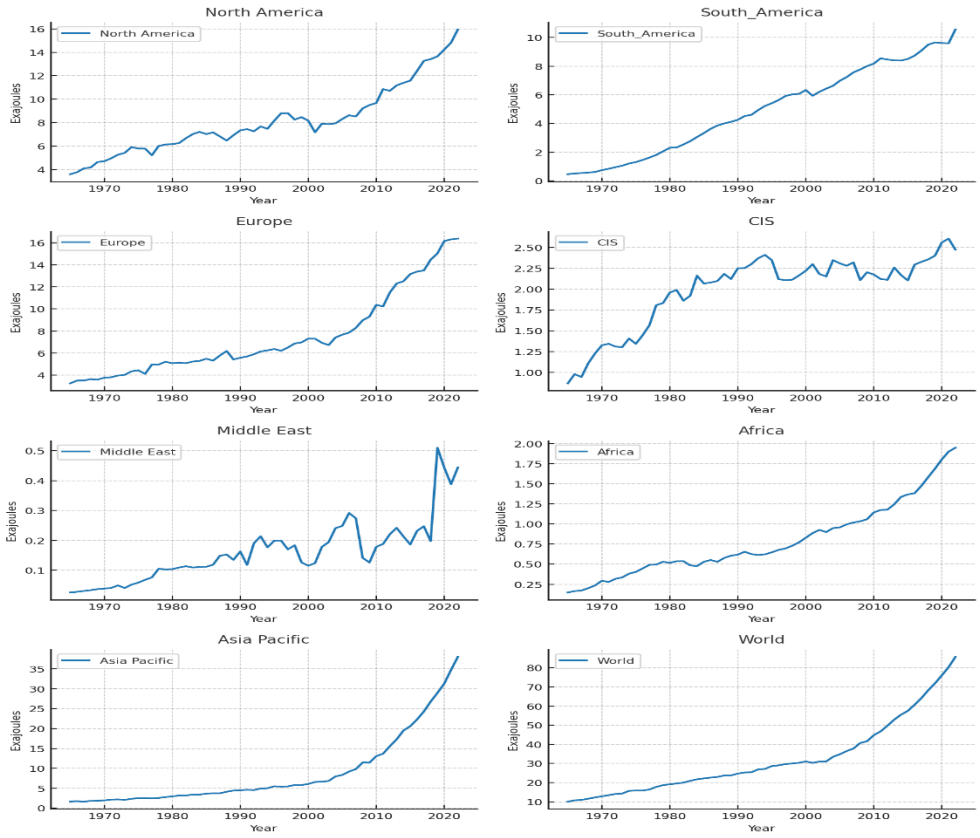


Figure 8: Annual trends in renewable energy consumption by region from 1965-2022

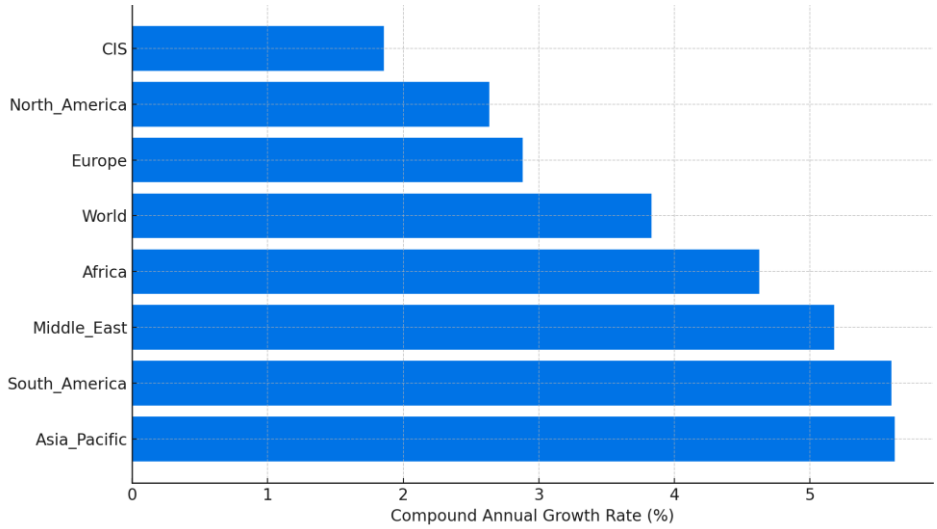


Figure 9: Ranking of Regions by Compound Annual Growth Rate.

Figure 8 shows that North America and Europe experienced a linear and stable growth in renewable energy consumption, while Asia-Pacific underwent an almost exponential increase starting in the early 2000s. South America also displays a strong linear growth over the study period. The CIS region recorded a sharp increase between 1965 and 1990, followed by a decline and near stabilization from the 1990s onward, likely due to the collapse of the USSR and the reclassification of some of its former countries into the European region. Africa and the Middle East exhibit moderate but noticeable increases in consumption over the 1965–2022 period, with a small surge observed in the late 2010s for both regions.

An analysis of Figure 9 reveals that Asia-Pacific and South America recorded the highest compound annual growth rates (CAGR) in renewable energy consumption over the study period, with 5,63% and 5,6% respectively. While North America and Europe maintain high absolute levels of renewable energy consumption, their annual growth rates are comparatively lower and more stable (2,88% and 2,63%, respectively). The CIS region exhibited the lowest growth among all regions, with a CAGR of 1,86%.Significant growth potential is observed in Africa (4,62%) and the Middle East (5,18%), both of which have nearly doubled their green energy consumption since the 2010s.

Figure 10 presents the evolution of fossil energy consumption by region over the period 1965–2022, while Figure 11 shows the average annual growth rate for the same period.

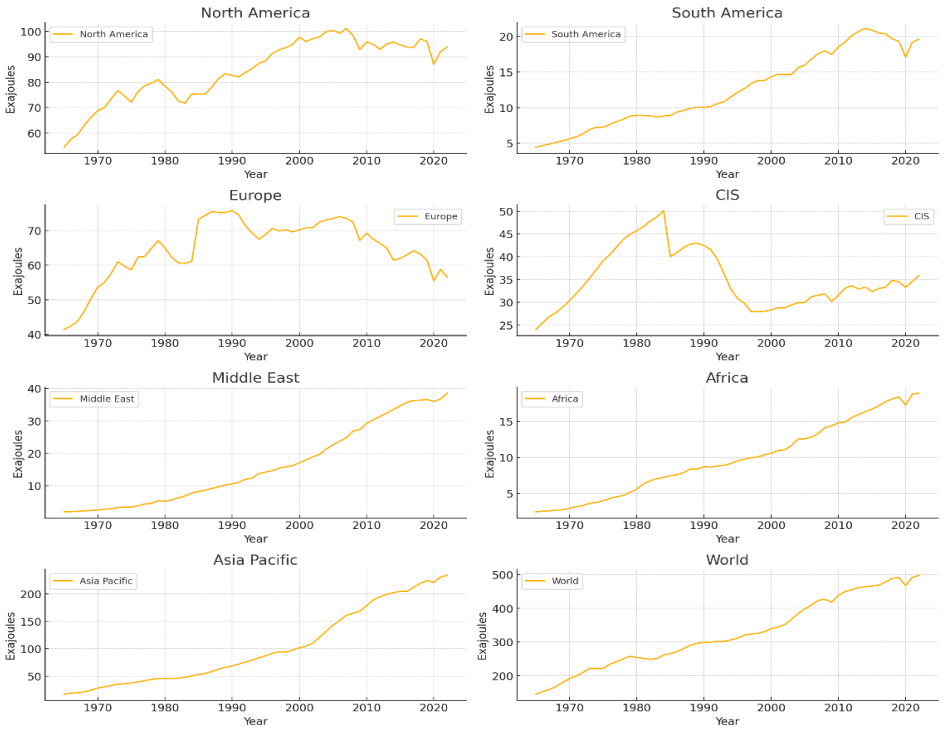


Figure 10: Annual trends in fossil energy consumption by region from 1965-2022

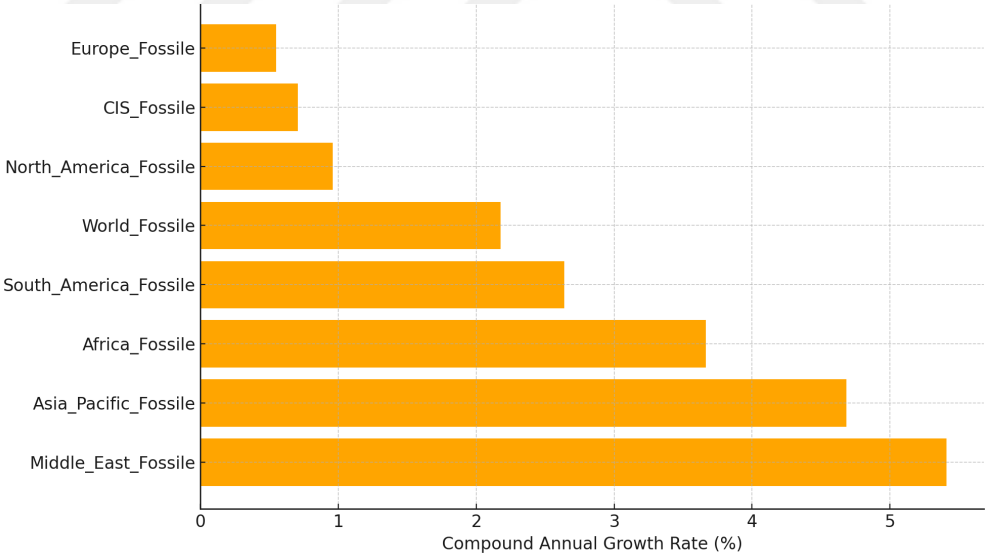


Figure 11: Ranking of Regions by Average Annual Growth Rate

After analyzing Figure 10 and Figure 11, it becomes evident that fossil energy consumption over the period 1965–2022 has generally increased, showing an almost linear trend in regions such as the Middle East, Asia-Pacific, Africa, and South America. These regions exhibit sustained and

steady growth, with respective average annual growth rates of 5,41%, 4,68%, 3.66%, and 2,67%. In contrast, North America (0,96% annually on average) saw consumption increase until the early 2000s, after which it began to stabilize. A similar trend is observed in Europe (0,55% annually), though with a general decline starting in the mid-2000s. The CIS region (0,71% annually) experienced rapid growth until the mid-1980s, followed by a sharp decline through the end of the 1990s, likely due to the collapse of the USSR. Nonetheless, a slow recovery in consumption has been observed since the early 2000s.

Overall, the analysis of the graphs reveals that renewable energy consumption has increased consistently across all world regions during the 1965-2022 period. The same applies to fossil fuel consumption, although in some regions, growth has continued uninterrupted, while in others, it has slowed, declined, or stabilized after a certain point.

In all series, we clearly observe a trend component and non-linearity, whereas no seasonal component is apparent. This observation suggests that the time series are non-stationary. This assumption was confirmed by the Augmented Dickey-Fuller (ADF) test at the 5% significance level, with results summarized in Table 4.

Table 4: ADF results

Data	***Series***	***ADF Statistic***	***p-value***
Renewable Data	*North_America_Renew*	2,058	1,00
	South_America_Renew	0,044	0,96
	Europe_Renew	3,07	1,00
	CIS_Renew	-2,52	0,11
	Middle_East_Renew	-0,14	0,95
	Africa_Renew	4,56	1,00
	Asia_Pacific_Renew	10,37	1,00
	World_Renew	2,95	1,00
Fossile Data	*North_America_ Fossile*	-2,76	0,065
	South_America_ Fossile	-0,72	0,84
	Europe_ Fossile	-2,74	0,067
	CIS_ Fossile	-1,79	0,39
	Middle_East_ Fossile	1,55	1,00
	Africa_ Fossile	1,61	1,00
	Asia_Pacific_ Fossile	2,12	1,00
	World_Fossile	-0,46	0,9

If we are to assess the stationarity of the time series, differencing operations must be applied (see Equation (8)). The order of differencing that successfully rendered each series stationary is recorded in Table 5.

Table 5: Order of differencing leading to stationarity.

Data	***Series***	***Differenciation Order***	***ADF Statistic***	***p-value***
Renewable Data	*North_America_Renew*	1	-7,10	4,09e-10*
	South_America_Renew	2	-3,99	1,43e-03**
	Europe_Renew	1	-4,02	1,33e-03*
	CIS_Renew	1	-7,77	8,76e-12*
	Middle_East_Renew	1	-4,54	1,65e-04*
	Africa_Renew	2	-7,37	8,80e-11**
	Asia_Pacific_Renew	3	-5,99	1,72e-07***
	World_Renew	2	-16,63	1,67e-29**
Fossile Data	*North_America_ Fossile*	1	-3,35	0,01*
	South_America_ Fossile	3	-4,64	1,07e-04***
	Europe_ Fossile	1	-6,12	9,01e-08*
	CIS_ Fossile	1	-5,88	3,04e-07*
	Middle_East_ Fossile	2	-6,69	4,01e-09**
	Africa_ Fossile	3	-5,94	2,31e-07***
	Asia_Pacific_ Fossile	2	-3,26	0,017**
	World_Fossile	3	-5,19	9,078e-06***

We observe that the orders of differencing required to achieve stationarity vary from one series to another. If we were to conduct our modeling using only stationary series, it would be necessary to adopt the highest order of differencing, i.e., 3 in our case. This would result in the loss of 3 years of data (representing approximately 5% of the dataset).

However, unlike classical statistical models such as ARIMA, neural network architectures like LSTM and their variants, particularly the A-LSTM model does not require time series to be stationary (Smyl, 2020; Bandara et al., 2020). These models are inherently capable of automatically

capturing trends, cycles, and complex non-linear behaviors in the data without the need for prior transformations such as differencing.

In the context of this thesis, it was decided to retain the time series in their original non-stationary form in order to:

- preserve the true dynamic structure of the data, particularly the long-term trends that are crucial in energy-related studies;
- fully leverage the deep learning capabilities of LSTM and A-LSTM models, which are specifically designed to handle such raw, non-stationary data;
- avoid information loss due to differencing, which could reduce the interpretability of the results or alter the relative contribution of certain variables.

Thus, the preprocessing was limited to data normalization, aimed at ensuring a stable convergence during model training, without applying any transformations to enforce stationarity. This methodological choice is supported by recent studies in the literature, which show that LSTM networks can produce excellent results when applied to non-stationary time series, particularly in the energy, economic, and financial domains.

3.2.1.4 Data Normalization

It was essential to normalize the data before integrating it into the A-LSTM model training process. This is because A-LSTM networks rely on activation functions such as tanh and sigmoid, which are sensitive to differences in value ranges. For instance, if one series has values between 0 and 1 while another spans from 0 to 100, it would create imbalances in the gradient calculations during backpropagation. Normalization brings all numerical values from different time series onto a common scale, with the goal of stabilizing the learning process and accelerating convergence.

To perform this step, the *MinMaxScaler*(.) function from Python was applied. The normalized results are presented in Figure 12.

Renewable Data Scaled

Year	North America	South_America	Europe	CIS	Middle East	Africa	Asia Pacific	World
1965-01-01	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.001005	0.000000
1966-01-01	0.013759	0.005026	0.019084	0.063594	0.004392	0.009488	0.002825	0.008809
1967-01-01	0.040644	0.007897	0.019339	0.043687	0.010591	0.013513	0.000000	0.011930
1968-01-01	0.046184	0.011216	0.029318	0.138377	0.016126	0.029763	0.004544	0.019786
1969-01-01	0.084034	0.015613	0.025420	0.206574	0.023029	0.048466	0.006284	0.028744

Fossil Data Scaled

Year	North_America_Fossile	South_America_Fossile	Europe_Fossile	CIS_Fossile	Middle_East_Fossile	Africa_Fossile	Asia_Pacific_Fossile	World_Fossile
1965-01-01	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1966-01-01	0.066336	0.014723	0.030141	0.056349	0.002049	0.006349	0.008068	0.022128
1967-01-01	0.105306	0.025001	0.068164	0.109356	0.005103	0.007614	0.012325	0.038442
1968-01-01	0.184784	0.042033	0.155038	0.142637	0.008276	0.015158	0.018825	0.065460
1969-01-01	0.251350	0.054207	0.260868	0.189136	0.011805	0.017936	0.034130	0.098605

Figure 12: Overview of the data set after normalization

3.2.2 Data Restructuring for Model Input

3.2.2.1 Creation of Time Series Sequences

Following the normalization step and in order to adapt the time series to the requirements of the A-LSTM model, the data were reorganized into sliding temporal sequences. This method is based on the so-called sliding window approach, where each observation fed into the model corresponds to a fixed-length multivariate sequence, known as a time step.

Considering our raw series $X = (X^1, X^2, \dots, X^n)^T = (X_1, X_2, \dots, X_m) \in \mathbb{R}^{(n*m)}$ and $X_t = (x_t^1, x_t^2, \dots, x_t^n) \in \mathbb{R}^n$ represents the feature vector of all regions at time t. $Y_t \in \mathbb{R}^m$,our output.

- $n = 7$ denotes the number of variables (or regions);
- $m = 58$ denotes the number of years (1965-2022).

To model temporal dependencies, we define a time window size (or time step) of $s = 4$. This means that, at each iteration, the model receives a sequence of 4 consecutive years as input to predict the value of the output variable at the year immediately following this sequence (5^{th} year). This division generates (2022 - 1965 + 1 - s) = 54 training samples. Here we cannot predict $y_{1965}, y_{1966}, y_{1967}, y_{1968}$ because there is no sufficient inputs which can predict them.

Given the time step $s = 4$, the construction of training sequences proceeds as follows:

$$(X_{t-3}, X_{t-2}, X_{t-1}, X_t) \xrightarrow{Predict} y_{t+1} \tag{35}$$

I.e.,

$$\textit{Sequence 1} : (X_{1965}, X_{1966}, X_{1967}, X_{1968}) \xrightarrow{Predict} y_{1969}$$
$$\textit{Sequence 2} : (X_{1966}, X_{1967}, X_{1968}, X_{1969}) \xrightarrow{Predict} y_{1970}$$
$$\textit{Sequence 3} : (X_{1967}, X_{1968}, X_{1969}, X_{1970}) \xrightarrow{Predict} y_{1971}$$
$$\vdots$$
$$\textit{Sequence 54} : (X_{2018}, X_{2019}, X_{2020}, X_{2021}) \xrightarrow{Predict} y_{2022}$$

3.2.2.2 Data Splitting

The training process was preceded by partitioning of the structured datasets. To assess the generalization ability of the A-LSTM model, the full dataset was chronologically divided into two subsets: a training set and a testing set. Given the temporal nature of the data, a non-random, time-respecting split was applied. Specifically, after generating the sequence pairs, the first 80% of the sequences were allocated for training the model, while the remaining 20% were reserved for evaluation (train_split: 80%; test_split: 20%).

Consequently, for a total of 54 temporal sequences in this study, the dataset was divided as follows:

$$(X_{train}, \mathrm{Y_{train}}) = (43, 43) \tag{36}$$

$$(X_{test}, \mathrm{Y_{test}}) = (11, 11) \tag{37}$$

This corresponds to 43 sequences for training and 11 sequences for testing.

This approach ensures that no future data leaks into the training process, thereby preserving the causal structure and the realism of the forecasting task.

3.2.3 Training Process

After preprocessing and data splitting, the datasets were passed to the proposed model for training. The model employed is an LSTM enhanced with an attention mechanism (A-LSTM) (see the architecture diagram).

Specifically, the architecture consists of a single LSTM layer that processes multivariate sequential data derived from time series. The second layer applies a dropout technique, used here to improve the model's generalization capacity and reduce overfitting. The third layer integrates the attention mechanism, allowing the network to learn to assign different weights to past time steps depending on their actual contribution to future prediction. Each intermediate output of the LSTM

is transformed into an attention score, which is then normalized using a softmax function. All time-step outputs are then combined in a weighted manner to form a context vector, a condensed representation of the most relevant information extracted from the input sequence.This vector is finally passed through a Dense layer to produce a prediction of global energy consumption at the current time step.

The final model is thus ready to be used directly for making predictions on the test data.

Figure 13 provides a summary of the training algorithm.

```
Input:
  - X_train, y_train: training data and labels
  - X_test, y_test: validation data and labels
  - n_steps: sequence length

Steps:
  1. input_seq ← Input(shape = (n_steps, dim_features))
  2. lstm_out ← LSTM(32, return_sequences = True)(input_seq)
  3. lstm_out ← Dropout(0.2)(lstm_out)

  4. attention_scores ← Dense(1, activation = 'tanh')(lstm_out)
  5. attention_weights ← Softmax(attention_scores, axis = 1)
  6. context ← weighted sum(attention_weights × lstm_out)

  7. output ← Dense(1)(context)
  8. model ← Model(inputs = input_seq, outputs = output)
  9. model.compile(optimizer = 'adam', loss = 'mse')

  10. early_stop ← EarlyStopping(patience = 10, restore_best_weights =
True)
  11. checkpoint ← ModelCheckpoint("best_model.h5", monitor =
'val_loss', save_best_only = True)

  12. model.fit(X_train, y_train,
              validation_data = (X_test, y_test),
              epochs = 200,
              batch_size = 16,
              callbacks = [early_stop, checkpoint])
```

Figure 13: A-LSTM network architecture implemented.

3.2.4 Hyperparameters

A hyperparameter is an external parameter to a machine learning model that is manually set before training and determines how the algorithm functions. These hyperparameters are tuned to ensure faster convergence and to prevent overfitting during training. In this study, the optimal values of the tunable hyperparameters were determined using a genetic selection algorithm, which efficiently explores the search space to identify high-performing configurations.

Genetic selection algorithms are stochastic optimization methods inspired by the principles of natural evolution, notably the selection of the fittest individuals, gene crossover, and mutation. They are particularly useful in contexts where the search space is large, non-linear, or

poorly structured, allowing for the discovery of optimal or near-optimal solutions to complex problems. The process generally begins with an initial population of randomly generated individuals, each representing a candidate solution. A cost (or performance) function evaluates each individual. The best-performing individuals are selected, recombined through crossover, and subjected to mutation to produce a new generation, following a logic of progressive improvement. This iterative mechanism is repeated over multiple generations until convergence toward a satisfactory solution is achieved (Holland, 1992).

In our study, each individual in the population represents a unique combination of the model's hyperparameters, including the number of epochs, *batch_size*, loss function, number of LSTM units, dropout rate, optimizer, activation functions used in the attention layers, the *patience* parameter associated with the Early stopping mechanism, the *restore_best_weights* option, as well as the *mode* for monitoring the metric (minimization or maximization). At each generation, the model is trained with a given hyperparameter combination and evaluated based on its predictive performance (using MSE on a validation set). The best-performing configurations are selected, crossed, and mutated to generate new solutions, enabling efficient automation and optimization of the model's architecture and training parameters.

Table 6 summarizes the optimized hyperparameters used in the A-LSTM model. The optimal number of training iterations (epochs) was set to a maximum of 100 for renewable energy data and 200 for fossil energy data. The number of memory cells in the LSTM layer was fixed at 32 units. A Dropout layer with a rate of 0.2 was added to regularize the network by randomly disabling 20% of the neurons during training. This helps prevent the model from relying too heavily on specific neurons or connections in the LSTM layer, thereby reducing overfitting. The model was trained using batches of 16 sequences at a time (*batch_size* = 16). The loss function used for training was the mean squared error (MSE), which was also applied on the validation dataset as the validation loss (*Val_loss*). The optimization algorithm employed to update the weights and minimize the loss function was Adam, ensuring that learning steps are scale-invariant with respect to parameter gradients. To prevent overfitting and reduce training time, an EarlyStopping mechanism was incorporated, which retains the model weights that performed best on the validation set. This mechanism was configured with a patience of 10, meaning training stops if no improvement in validation loss is observed for 10 consecutive epochs.

Table 6: Optimized hyparameters for the A-LSTM model

Hyperparameters	Values for Renwables	Values for Fossiles
Epoch	100	200
Bacth_size	16	16
Loss fucntion	MSE	MSE
LSTM unity	32	32
Dropout	0,2	0,2
Optimizer	Adam	Adam
Attention activation functions	Tanh and sigmoid	Tanh and sigmoid
patience	10	10
restore_best_weights	True	True
Mode	minimum	Minimum

3.2.5 Prediction

After training and saving the best model, we performed predictions using the test dataset, which consists of data that had not been included during training. The input sequences were passed to the A-LSTM model via the *predict(.)* function, which generated normalized estimates of global energy consumption. These predictions were then converted back to their original scale using *scaler.inverse_transform (.)*, allowing direct comparison with the actual observed value.

Each prediction represents an estimate of global consumption for a given year, based on the historical data from the previous four years across the seven regions. This step enabled the evaluation of the model's real-world performance in a practical forecasting context.

3.3 Implementation Phase

3.3.1 Development Environment

The entire modeling process was carried out using the **Python** programming language (*version 3.10*), chosen for its extensive libraries dedicated to time series processing and deep learning. The development was conducted on the **Google Colab** platform, which offers a powerful and free cloud-based infrastructure, well-suited for training neural networks.

3.3.2 Libraries Used

The implementation of the A-LSTM network required several Python libraries, including:

- ✓ TensorFlow (along with its high-level API, Keras) for building, training, and evaluating the A-LSTM model;
- ✓ scikit-learn for data normalization (MinMaxScaler), data splitting (train_test_split), and performance evaluation metrics (MAE, RMSE, etc.);
- ✓ NumPy and Pandas for handling tabular structures and managing time series sequences;
- ✓ Matplotlib for plotting and visualizing the results.

All libraries were used in their stable and well-documented versions, with a strong emphasis on code reproducibility. The notebooks were organized in a modular fashion to allow easy reuse of preprocessing, training, and evaluation functions.

3.3.3 Learning Parameters

The training process of the model was configured using several key hyperparameters, including:

- ➢ **Batch size**: which defines the number of samples processed before the network's internal weights are updated;
- ➢ **Number of epochs**: representing the total number of complete passes through the training dataset. In this study, we opted for 50 epochs;
- ➢ **Dropout rate**: used as a regularization method to prevent overfitting;
- ➢ **Adam optimization algorithm**: selected for its robustness and dynamic adaptation of the learning rate during training;
- ➢ **EarlyStopping**: a regularization mechanism that halts training when no significant improvement is observed over a predefined number of consecutive epochs.

These parameters, specifically the number of hidden units, dropout rate, batch size, and number of epochs, were manually selected based on recommendations from the literature.

In summary, the model implementation followed the steps below:

(1) Setting up the Google Colab environment;

(2) Importing the necessary libraries: TensorFlow, Keras, NumPy, Matplotlib, Math, and scikit-learn;

(3) Importing the core network layers;

(4) Loading the datasets;
(5) Indexing the "Year" variable to serve as the chronological reference;

(6) Analyzing the characteristics of the time series;
(7) Normalizing the data, defining the sliding window size, and splitting the dataset into training and testing sets;
(8) Building the input layer using an LSTM network;
(9) Integrating the Attention layer and setting up the final output layer;
(10) Setting up the best model configuration and evaluating its performance;
(11) Running predictions;
(12) Saving the model.

3.4 Model evaluation metrics

The performance of the proposed model was evaluated using four metrics:

- **Mean Square Error (MSE):**

The MSE is the average of the squared deviations between predicted and actual values. The prediction objective is that its value should tend towards or be equal to 0. It is given by Equation (38) :

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{38}$$

- **Mean Absolute Error- MAE :**

It measures the average absolute error between predictions and observed values. It is given by the Equation (39):

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{39}$$

- **Root Mean Squared Error (RMSE)**

The RMSE measures the standard deviation of residuals. It's equal to square root of MSE. It's given by Equation (40) :

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \quad (40)$$

- **Mean Absolute Percentage Error (MAPE)**

The MAPE measures how close the values predicted by a model are to the actual values, in percentages. It's given by Equation (41):

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| * 100 \quad (41)$$

3.5 Conclusion

This chapter has provided a detailed account of the methodological steps undertaken to design, structure, and train an A-LSTM model for predicting global energy consumption based on regional data. The approach adopted follows a rigorous logic, combining standard time series processing practices (windowing, normalization, reshaping) with the use of advanced neural architectures integrating an attention mechanism.

Data preparation enabled the generation of multivariate time sequences compatible with the input format required by the model. The neural network implementation relied on robust and reliable tools such as TensorFlow and Keras in the Python environment, offering significant flexibility for experimentation and automatic hyperparameter optimization.

Finally, the training process was carefully calibrated through an internal cross-validation framework, allowing for the selection of the model with the best performance on unseen data. These methodological choices aim to ensure the robustness, reproducibility, and relevance of the results presented in chapter 4.

CHAPTER IV: Results and Discussion

4.1 Introduction

This chapter presents the empirical results obtained following the implementation of the A-LSTM model applied to our two datasets related to energy consumption. It also outlines the forecasting results with the aim of evaluating the predictive performance of the proposed model. In addition, we considered a comparative analysis with other models from the literature.

4.2 Presentation of Training Results

After establishing the architecture for training, the implementation process was carried out. Upon completion of the training, the results were obtained and are summarized in Figures 14 and 15. These figures illustrate the LSTM layers, with *Dense_2* representing the output layer to which the attention mechanism is applied. *Lambda_2* and *Lambda_3* correspond respectively to the layer that computes attention weights using the *softmax* function, and the layer that calculates the context vector within the mechanism.

The training process lasted approximately 29 seconds, involving a total of 15 560 parameters of which 5 186 were trainable parameters and 10 374 were associated with the optimizer.

Model: "functional_1"

Layer (type)	Output Shape	Param #	Connected to
input_layer_1 (InputLayer)	(None, 4, 7)	0	-
lstm_1 (LSTM)	(None, 4, 32)	5,120	input_layer_1[0]…
dropout_1 (Dropout)	(None, 4, 32)	0	lstm_1[0][0]
dense_2 (Dense)	(None, 4, 1)	33	dropout_1[0][0]
lambda_2 (Lambda)	(None, 4, 1)	0	dense_2[0][0]
lambda_3 (Lambda)	(None, 32)	0	dropout_1[0][0], lambda_2[0][0]
dense_3 (Dense)	(None, 1)	33	lambda_3[0][0]

Total params: 15,560 (60.79 KB)
Trainable params: 5,186 (20.26 KB)
Non-trainable params: 0 (0.00 B)
Optimizer params: 10,374 (40.53 KB)
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 48ms/step - loss: 1.9305e-04
Test MSE: 0.00019305205205455422

Figure 14: Summary of A-LSTM parameters after data learning on Renewable data.

The model applied to fossiles involved a total of 15 560 parameters, all of which were trainable, indicating a fully learnable architecture tailored to capture the underlying dynamics of the data.

Model: "functional_18"

Layer (type)	Output Shape	Param #	Connected to
input_layer_18 (InputLayer)	(None, 4, 7)	0	-
lstm_18 (LSTM)	(None, 4, 32)	5,120	input_layer_18[0…
dropout_18 (Dropout)	(None, 4, 32)	0	lstm_18[0][0]
dense_36 (Dense)	(None, 4, 1)	33	dropout_18[0][0]
lambda_36 (Lambda)	(None, 4, 1)	0	dense_36[0][0]
lambda_37 (Lambda)	(None, 32)	0	dropout_18[0][0], lambda_36[0][0]
dense_37 (Dense)	(None, 1)	33	lambda_37[0][0]

Total params: 15,560 (60.79 KB)
Trainable params: 5,186 (20.26 KB)
Non-trainable params: 0 (0.00 B)
Optimizer params: 10,374 (40.53 KB)
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 63ms/step - loss: 6.3361e-04
Test MSE: 0.0006336148944683373

Figure 15: Summary of A-LSTM parameters after data learning on Fossile data.

4.3 Loss Curve

After training, it was essential to analyze the evolution of the loss function of the A-LSTM model, measured on both the training set (Training Loss) and the validation set (Validation Loss), expressed as the MSE. This metric serves as a key indicator for evaluating the quality of the neural network's learning process and its generalization capability.

4.3.1 Loss Curve after renewable energy training

The loss curve associated with the training on renewable energy data (Figure 16) shows that the training loss remains consistently low and stable from the first epoch. It does not exhibit significant fluctuations, indicating that the model is capable of efficiently learning patterns present in the training data. The validation loss initially starts at a relatively high level (≈ 0.13) between epochs 0 and 10, reflecting a substantial gap between the model's predictions and the actual data. However, a rapid and steady decrease is observed from the first iterations. This downward trend

continues until around epoch 50, after which the curve stabilizes with a loss value approaching zero. This behavior reflects a progressive reduction of errors, suggesting increasingly accurate modeling of the underlying trends in the data.

The Best Epoch, identified here at epoch 80, corresponds to the point where the validation loss reaches its minimum value. This marks the optimal moment when the model provides the best generalization performance, prior to any potential overfitting.

The gap between the training and validation loss curves gradually narrows over time, which reflects a reduction in the generalization error. This indicates that the model is not merely memorizing the training data but is effectively learning generalizable representations.

In summary, the learning process of the A-LSTM model was fast, stable, and effective, demonstrating an excellent match between the network architecture and the nature of renewable energy data. The choice of an adequate number of epochs (100), combined with appropriate regularization (Dropout), ensured convergence toward a robust and generalizable model.

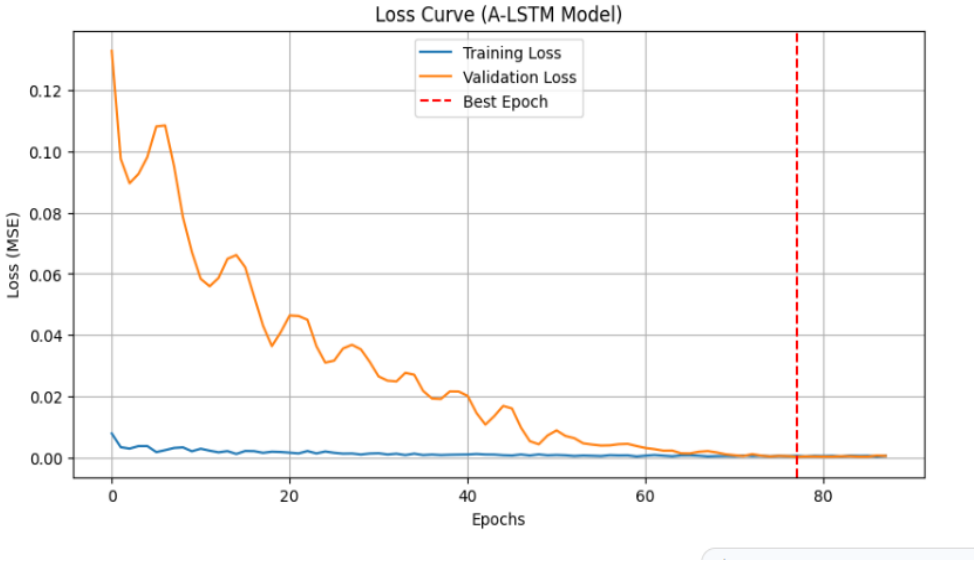


Figure 16: Loss Curve of A-LSTM applied on Renewable data.

4.3.2 Loss curve after fossil energy training

In contrast, the loss curve when the model is applied to fossil energy data (Figure 17) shows a faster and more pronounced convergence. The validation loss remained high between epoch 0 and epoch 5, before dropping rapidly and beginning to stabilize around epoch 10. From the 10^{th} epoch onward, the validation loss and training loss curves are closely aligned, suggesting that the model is well-regularized and does not suffer from learning bias or overfitting issues. No significant rebound is observed in the validation loss after the optimal point, which confirms the strong stability of the model. The Best Epoch was reached at epoch 17.

Consequently, the smooth and symmetric behavior of both curves confirms that the network was properly trained, with a well-calibrated parameter setup (dropout rate, batch size, learning rate). The A-LSTM model's fit to the data is both deep and well-controlled, as evidenced by the harmonious convergence of the training and validation curves from the earliest epochs. This reflects the model's strong capacity to accurately learn and generalize from the fossil energy data series.

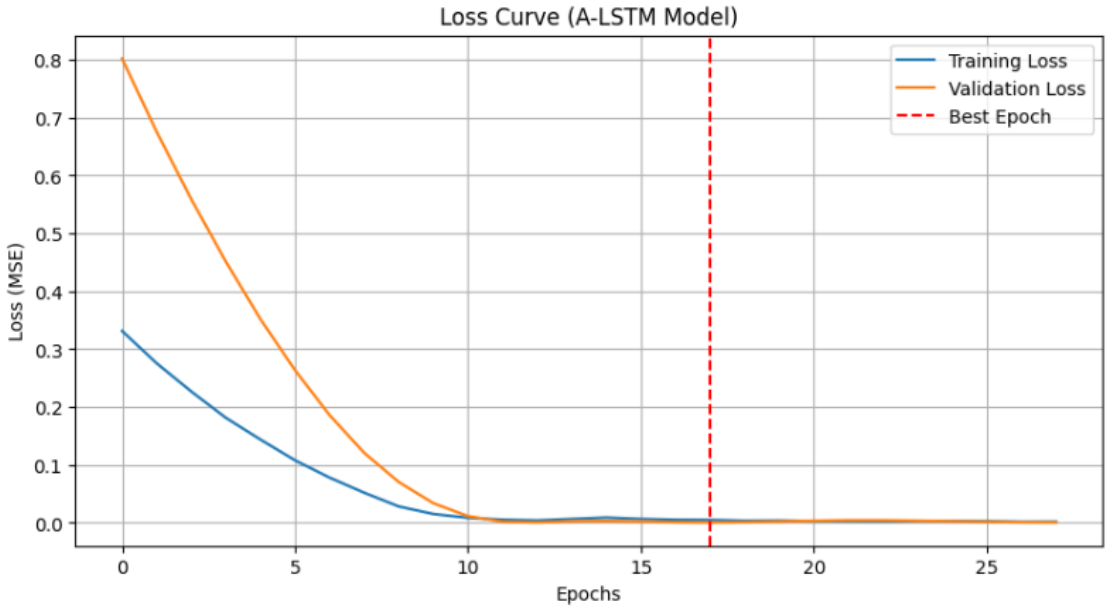


Figure 17: Loss Curve of A-LSTM applied on Fossile data

4.4 Prediction

At the end of the training phase, we proceeded with the prediction of the data. It is important to note that this prediction was performed on a test set consisting of 11 sequences, based on the split (80% for training and 20% for testing) carried out in Chapter 3.

4.4.1 Prediction of renewable energy consumption

Figure 18, which compares the actual and predicted values, shows an almost perfect overlap between the two trajectories over the test period. The overall shape of the curve, reflecting a steady increase in consumption, is accurately reproduced by the A-LSTM model. The curves evolve in parallel without significant deviation, indicating the model's strong ability to capture the underlying trend. This is further supported by Table 7, which shows that the absolute error remains below 1 unit for most predictions. The highest error recorded is 2,26 units and occurs at the beginning of the series (Year=2013), where the values are lower and therefore more sensitive in relative terms. The error tends to decrease as consumption increases, demonstrating the model's robustness at higher values. This distribution of errors highlights excellent predictive stability and

confirms that the model does more than simply follow an average trend; it successfully adapts locally to fluctuations.

The performance metrics obtained indicate a strong learning capacity of the model, with a MSE of 0,0002, which is extremely low (Table 8). The difference between the actual and predicted values remains minimal, as shown by an MAE of 0,0116. The MAPE of 1,77% indicates that the average deviation between actual and predicted values is very small, even at lower levels of consumption. Overall, the model demonstrates high robustness and effectively captures the key dynamics of the time series.

Thus, the results confirm the A-LSTM model's ability to learn the specific temporal patterns of global renewable energy consumption. Thanks to the integration of the attention mechanism, the model can prioritize the most relevant temporal information, thereby enhancing the accuracy of predictions which remains excellent throughout the test phase.

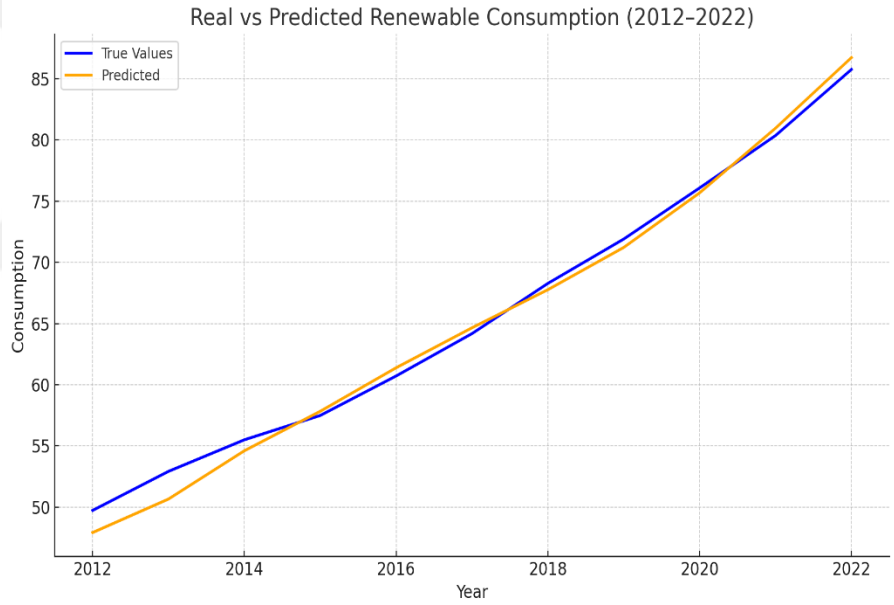


Figure 18: Evolution of actual vs. predicted renewable data

Table 7: Absolue Errors between actual and predicted of renewable values.

Year	Real	Predicted	Absolute Error
2012	49.737445	47.928577	1.808867
2013	52.927285	50.666954	2.260331
2014	55.505888	54.610828	0.895059
2015	57.494020	57.834896	0.340876
2016	60.719458	61.390011	0.670553

2017	64.181254	64.656548	0.475294
2018	68.282679	67.768784	0.513895
2019	71.907347	71.221855	0.685492
2020	76.083844	75.680885	0.402959
2021	80.360538	80.960541	0.600003
2022	85.757355	86.725571	0.968216

Table 8: Metrics results_Renewable

Metrics	Values
MSE	0,0002
MAE	0,0116
RMSE	0,014
MAPE (%)	1,17

4.4.2 Prediction of fossil energy consumption

Regarding the comparison between actual and predicted values on the test set for fossil energy consumption, Figure 19 shows an overall satisfactory alignment, although certain fluctuations in the real series are not entirely captured by the model's predictions. The model accurately follows the upward trajectory of consumption up to year 7 (corresponding to 2018). A significant divergence is observed around t=8 (2020), where actual consumption drops sharply, while the prediction continues along a linear trend. This highlights the model's difficulty in anticipating sudden or atypical changes. From t=9 (2021) onward, the model adjusts its estimation but remains slightly behind the actual rebound in values.

Table 9, which presents prediction errors, reveals more variability than in the renewable energy dataset. Most errors fall between 1 and 7 units, which remains moderate given the absolute value range of the data (450-500). Two major error peaks are observed at t=2013 (MAE=10,47) and t=2020 (MAE=21,83). This table indicates that the largest deviations occur in the most recent years, which may be attributed to the increasing complexity of recent dynamics in fossil energy consumption. The large discrepancy at t = 2020 is likely related to an external shock; most probably the COVID-19 pandemic, which caused a significant drop in real energy demand.

Despite this, the model achieved noteworthy performance, with the following evaluation metrics: MSE=0,0006, MAE=0,019, RMSE=0,027, and MAPE=2,07% (see Table 10). These

metrics indicate a high level of accuracy, with a very low average relative error, below 3%. This level of performance is particularly remarkable given the more volatile and irregular nature of fossil energy consumption time series. The MAPE of 2,07% confirms that the model remains reliable even in the presence of large value amplitudes.

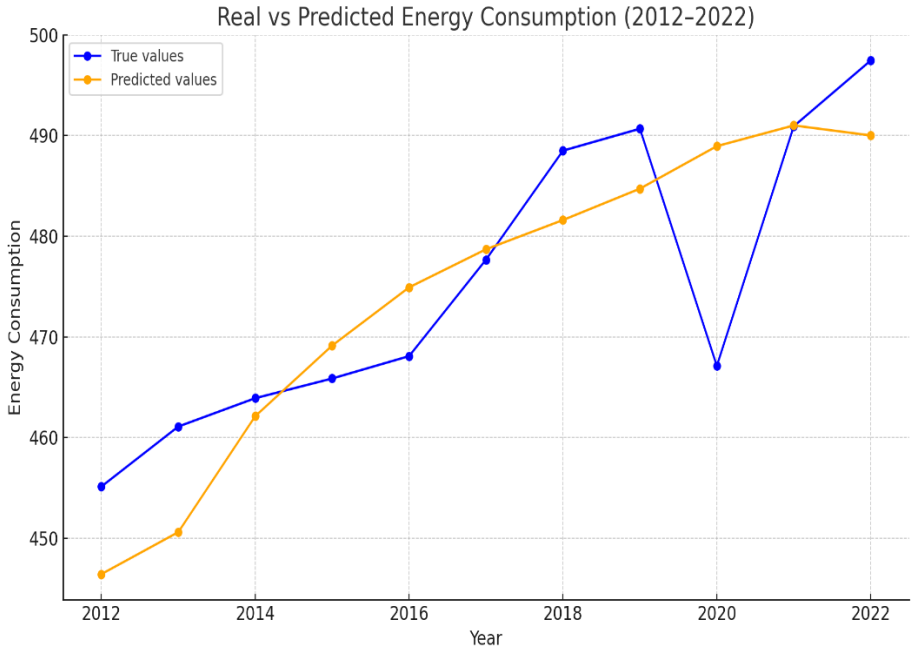


Figure 19: Evolution of actual vs. predicted Fossil data

Table 9: Absolue Errors between actual and predicted of fossil values.

Year	Real	Predicted	Absolute Error
2012	455.142485	446.447296	8.695189
2013	461.103789	450.616913	10.486876
2014	463.920121	462.131744	1.788376
2015	465.875383	469.147980	3.272597
2016	468.097425	474.928009	6.830584
2017	477.682774	478.709137	1.026363
2018	488.504245	481.613312	6.890933
2019	490.681970	484.728424	5.953546
2020	467.123268	488.952332	21.829064
2021	490.930539	491.010223	0.079685
2022	497.458227	490.017365	7.440863

Table 10: Metric results_Fossile

Metrics	Values
MSE	0,0006
MAE	0,019
RMSE	0,025
MAPE (%)	2,07

Overall, the results obtained from the A-LSTM model highlight the robustness, accuracy, and adaptability of our architecture in modeling global energy consumption whether renewable or fossil.

For renewable energy data, the model demonstrates outstanding performance, both in terms of error metrics (MAPE = 1,77%, MAE = 0,0116) and the consistency of its predictions. The smooth, gradual, and continuous temporal structure of renewable consumption is particularly well captured by the LSTM combined with the attention mechanism. This allows the model to reproduce the general trend with near-perfect precision. The forecast curves closely match the actual observations, and errors remain consistently low, stable, and without spikes.

In contrast, for fossil energy data, the model encounters more difficulty in predicting sudden fluctuations. Although the overall performance remains very good (MAPE = 2,07%, MAE = 0,0192), certain specific instances show larger errors, especially during abrupt trend reversals.

Despite these localized discrepancies, the A-LSTM model demonstrates strong generalization capabilities, mastery of complex temporal structures, and reliable performance in both energy contexts. Furthermore, a comparison was made between our A-LSTM model and other models from literature like the MLP, ARIMA-ANN, and standard LSTM models previously discussed in Chapter 2.

4.5 Comparison of the Model with Other Approaches

4.5.1 Description of the Models

We initiated a comparison of our A-LSTM model with three other models frequently cited in the literature: ARIMA-ANN, MLP, and the standard LSTM. This comparison was carried out using rigorous and consistent criteria, ensuring that all models operated under the same experimental conditions.

We begin with the MLP model, which like most Artificial Neural Networks, does not require stationnarity of the data. The raw data were first normalized, then split into training and test sets using an 80%-20% chronological split. Since MLP inputs do not require sequential data structures, the 58 available observations were divided into 46 training samples (X_train and Y_train) and 12 test samples (X_test and Y_test). We defined an MLP model with a single hidden layer of 32 units and a batch size of 16 for both renewable and fossil datasets. The difference lies in the training epochs: 100 epochs for renewable energy and 200 epochs for fossil fuels, both using an *EarlyStopping* mechanism to prevent overfitting.

Next, the ARIMA-ANN model, as discussed in Chapter 2, is a hybrid model that uses the residuals from an ARIMA model as inputs for an artificial neural network. Since ARIMA requires stationary time series, we first stationarized the data. Stationarity was achieved through third-order differencing for the global renewable consumption series (World_Renew) and second-order differencing for the global fossil consumption series (World_Fossil), as detailed in table 3 (Chapter 3). The ARIMA model parameters (p, d, q) were set as (2, 3, 2) for renewable energy and (2, 2, 2) for fossil energy. The values of parameters p and q were determined after several attempts by considering the model that gave the lowest AIC criterion (Akaike Information Criterion). After fitting the ARIMA model, we extracted the residuals and used them to train a neural network. As with the MLP and A-LSTM, the ANN was defined with a single layer of 32 units, a batch size of 16, and 100 and 200 epochs for renewables and fossils respectively.

Finally, the LSTM model used here is the standard version without attention mechanism, similar to the base of the proposed A-LSTM model. It consists of a single LSTM layer with 32 units, a batch size of 16, and 100 and 200 epochs respectively for the renewable and fossil datasets. The time step was set to 4, which resulted in 43 sequences for training and 11 sequences for testing in each case.

4.5.2 Training outcome

The performance of the four models was assessed separately on the two datasets. The results obtained using evaluation metrics such as MSE, MAE, RMSE, and MAPE (%) provide a comparative analysis of model behavior based on the nature of the time series involved. These results are summarized in Tables 11 and 12.

The output reveals significant differences in prediction accuracy across models but also depending on the nature of the time series. The MLP model consistently showed the poorest performance in both cases. For renewable energy data, it reached a MAPE of 12,76%, while it slightly improved on fossil fuel data with a MAPE of 7,91%. This suggests that the MLP model struggles to learn the temporal dynamics and structural patterns of our energy datasets.

The hybrid ARIMA-ANN model, on the other hand, produced noticeably better results, particularly in reducing absolute error for both data types. It achieved a MAPE of 5,92% on fossil fuels, compared to 12,76% on renewables, indicating stronger performance on more stable or structured series.

In contrast, the standard LSTM model demonstrated remarkable robustness, outperforming both the MLP and ARIMA-ANN models. It achieved a MAPE of 2,66% on renewables and 2,31% on fossil fuels, showing its capacity to learn time-dependent patterns even in non-linear series.

Finally, the proposed A-LSTM model, enhanced with an attention mechanism, outperformed all other models, delivering optimal results on both datasets. It recorded a MAPE of 1,77% for renewable energy and 2,07% for fossil fuels, outperforming the standard LSTM even further. This is a noteworthy finding, especially considering the well-established effectiveness of the LSTM architecture in modeling temporal dependencies in time series forecasting.

Table 11: Metrics results for different models using renewable datasets.

	Model	**MSE**	**MAE**	**RMSE**	**MAPE (%)**
Renewable	MLP	0,011	0,094	0,1	12,76
	ARIMA-ANN	0,002	0,048	0,039	12,76
	LSTM	0,0005	0,018	0,022	2,66
	A-LSTM proposed	**0,0002**	**0,0116**	**0,014**	**1,77**

Table 12: Metrics results for different models using fossil dataset.

	Model	**MSE**	**MAE**	**RMSE**	**MAPE (%)**
Fossile	MLP	0,0099	0,0756	0,099	7,91
	ARIMA-ANN	0,0034	0,028	0,058	5,92
	LSTM	0,0007	0,022	0,027	2,31
	A-LSTM proposed	**0,0006**	**0,019**	**0,025**	**2,07**

In summary, traditional models such as MLP and ARIMA-ANN demonstrate a certain degree of fragility when switching between different types of energy data, whereas memory-based models like LSTM, and especially A-LSTM, offer robust and consistent performance. This overall comparison highlights the relevance of using recurrent architectures with attention mechanisms for high-value energy forecasting applications.

4.6 General Discussion

This study focused on forecasting energy time series data using the A-LSTM model, a relatively recent addition to the literature. The choice of this model proved particularly compelling given the predictive results obtained. By combining the memory capabilities of the LSTM with an attention mechanism, the A-LSTM offers a more refined approach to temporal analysis. Unlike traditional models, it does not treat the entire past sequence uniformly, but instead learns to focus on the periods that have the most significant impact on the target variable.

The comparative analysis of the proposed A-LSTM model with other models such as MLP, ARIMA-ANN, and the standard LSTM, based on evaluation metrics like MSE, MAE, and MAPE, clearly showed that the A-LSTM delivered the best predictive performance, both for renewable and fossil energy consumption. These results demonstrate that attention-based architecture is not only more flexible, but also more effective in modeling complex and sometimes irregular dynamics, especially considering the nonlinear nature of the energy data analyzed in this study.

This research supports previous studies that have applied A-LSTM in other domains. Ran et al. (2019) found that the model outperformed SARIMA, linear regression, Recurrent State-space Neural Networks (SSNN), and even standard LSTM in travel time prediction. Similarly, Qiu et al. (2020), Suang and Li (2024), and Sebastian and Tantia (2025) also achieved better results with LSTM combined with attention mechanisms in stock market forecasting.

4.6.1 Projection of Global Energy Dynamics by 2026

Based on the proposed A-LSTM model, forecasts were carried out not only on the test dataset composed of 11 sequences but also extended to cover the period 2023-2026. The analysis of the results revealed interesting trends for both types of energy sources.

4.6.1.1 Renewable Energy

According to the model's forecasts, global consumption of renewable energy shows a steady increase, rising from 47,92 exajoules in 2012 to 86,92 exajoules in 2022. This consumption is further projected to grow from 93,15 exajoules in 2023 to 101,83 exajoules in 2026 (Figure 20). The average annual growth rate (CAGR) over this period is estimated at 5,73%, indicating strong

growth momentum. This trend likely reflects the cumulative effects of climate policies, technological investments, and structural changes in national energy mixes. If this trajectory

continues, renewable energies could play an increasingly central role in meeting global energy demand in the medium term.

4.6.1.2 Fossil Energy

In contrast, global fossil energy consumption is expected to evolve more moderately over the same period. Projections indicate a progression from 488,45 exajoules in 2023 to 497,91 exajoules in 2026, representing an average annual growth rate of just 0.02%, which is extremely low or nearly stagnant (Figure 20). This result suggests a possible saturation in demand, potentially linked to energy efficiency efforts, the transition to decarbonized sources, or geopolitical factors limiting fossil fuel expansion. It could represent a turning point, signaling the end of structural growth in fossil energy, even though its absolute consumption levels remain high.

Figure 20 below presents a comparative analysis of the model's predictions for both energy types over the 2022-2026 period as generated by the A-LSTM model.

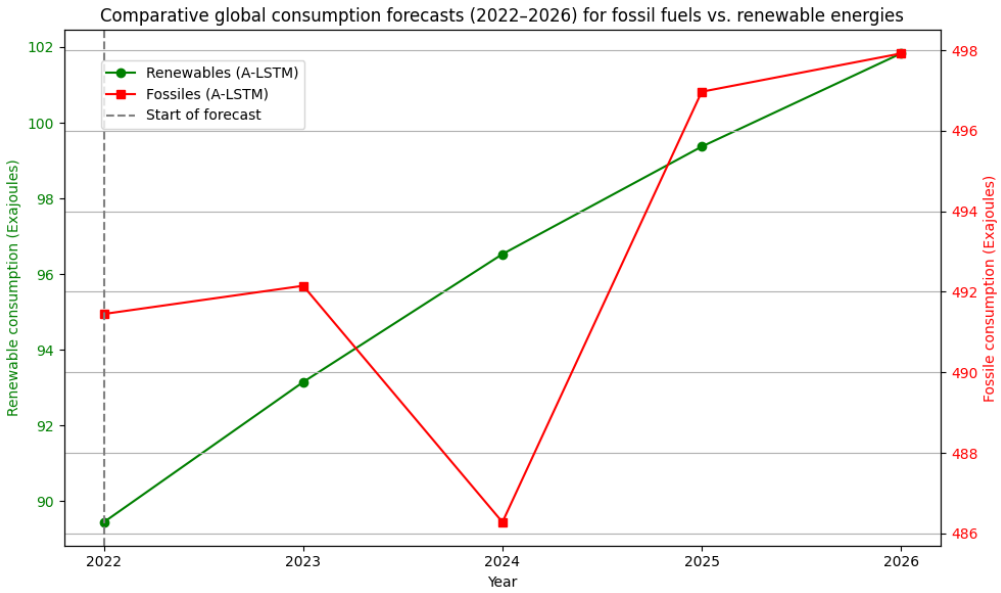


Figure 20: Comparative gloabal consumption forcasts (2022-2026) for fossil fuels vs renewables energies.

The gap between the two trajectories reflects an energy bifurcation: on one hand, the continuous rise of renewables; on the other, the stagnation or even a transition toward decline for fossil fuels. This dual dynamic lies at the heart of scenarios compatible with the objectives of the Paris Agreement and the 2030 commitments. However, to confirm this structural shift, it will be

necessary to monitor developments beyond 2026 and determine whether the trend persists or reverses.

4.6.1.3 Limitations of Our Study

- **Dataset size**: Although the model produced good results on our datasets, it should be noted that their size was relatively small (58 observations). We would have preferred to work with longer and broader datasets to enhance the model's robustness.
- **Lack of data on exogenous factors**: Including exogenous variables in the model could help improve the accuracy of predictions—especially for fossil energy consumption, which is significantly influenced by external factors such as stock prices, oil prices, and other macroeconomic indicators.

4.7 Conclusion

The results obtained using the A-LSTM model demonstrated superior forecasting performance, achieving a MAPE of 1,77% for renewable energies and 2,07% for fossil fuels, thereby outperforming the MLP, LSTM, and ARIMA-ANN models. The extended forecasts for the period 2023-2026 reveal a steady growth in renewable energy consumption (average annual growth rate = 5,73%) compared to an almost stagnant trend for fossil fuels (average annual growth rate = 0,02%). These contrasting trajectories underscore an ongoing energy transition, clearly captured by the accuracy and adaptability of the proposed model.

GENERAL CONLUSION

The objective of this study was to evaluate, through a predictive approach, the trajectory of global fossil and renewable energy consumption, in order to assess the potential for an energy transition by 2030. To this end, we implemented an A-LSTM (Attention-based Long Short-Term Memory) model, known for its effectiveness in processing multivariate time series and its ability to weigh the influence of past sequences. The model was applied to two datasets comprising seven input variables representing energy consumption across major global regions (CIS, North America, South America, Europe, Africa, Asia-Pacific, and the Middle East), and one output variable corresponding to global energy consumption, over the period 1965–2022.

The results showed that the A-LSTM model was capable of generalizing and tracking the trends in both datasets with very low error metrics. Forecasting performance was excellent, and comparative analyses with other models such as MLP, ARIMA-ANN, and the standard LSTM revealed that the proposed A-LSTM outperformed them all. This highlights the robustness and relevance of the A-LSTM architecture in energy forecasting.

Furthermore, in line with our research goals, we extended the prediction horizon beyond the observed data up to the year 2026. The forecasted results highlighted two contrasting dynamics: on the one hand, fossil energy consumption appears to plateau over the 2023-2026 period, with an almost flat average annual growth rate (0,2%), indicating a potential saturation in global demand. On the other hand, renewable energy consumption shows steady growth at an average annual rate of 5,73%, reflecting a dynamic consistency with global energy transition goals.

However, despite this upward trend, fossil fuels continue to dominate in absolute consumption volumes, underscoring the slow structural shift in the global energy landscape. Therefore, while the results point toward a favorable evolution in the share of renewables, their capacity to significantly replace fossil fuels in the short term remains uncertain. This situation calls for intensified political, economic, and technological efforts to accelerate the transition, especially in sectors with high carbon inertia.

In the future, this work could be expanded by incorporating exogenous variables (such as energy prices, investments, public policies, etc.), extending the forecasting horizon, or comparing with other advanced deep learning architectures. Such developments would enhance the robustness of predictions and provide deeper insights for strategic decision-making in global energy policy.

REFERECENCES

Abas, N., Kalair, A., and Khan, N., 2015. Review of fossil fuels and future energy technologies. Futures, 69, 31–49. https://doi.org/10.1016/j.futures.2015.03.003

Alex., 2024. Global Wind Report 2024. Global Wind Energy Council. https://gwec.net/global-wind-report-2024/

Anderson, O. D., 1977. The Box-Jenkins approach to time series analysis. RAIRO-Operations Research, 11(1), 3–29.

Archer, R., 2020. 20—Geothermal Energy. In T. M. Letcher (Ed.), Future Energy (Third Edition) (pp. 431–445). Elsevier. https://doi.org/10.1016/B978-0-08-102886-5.00020-7

Bahdanau, D., Cho, K., & Bengio, Y., 2014. Neural machine translation by jointly learning to align and translate. arXiv Preprint arXiv:1409.0473. https://peerj.com/articles/cs-2607/code.zip

Bandara, K., Bergmeir, C., & Hewamalage, H., 2020. LSTM-MSNet: Leveraging forecasts on sets of related time series with multiple seasonal patterns. IEEE transactions on neural networks and learning systems, 32(4), 1586-1599.

Barbier, E., 2002. Geothermal energy technology and current status: An overview. Renewable and Sustainable Energy Reviews, 6(1), 3–65. https://doi.org/10.1016/S1364-0321(02)00002-3

Benchaji, I., Douzi, S., El Ouahidi, B., & Jaafari, J., 2021. Enhanced credit card fraud detection based on attention mechanism and LSTM deep model. Journal of Big Data, 8(1), 151. https://doi.org/10.1186/s40537-021-00541-8

Blaabjerg, F., & Ma, K., 2017. Wind energy systems. Proceedings of the IEEE, 105(11), 2116–2131.

Bre, F., Gimenez, J. M., & Fachinotti, V. D., 2018. Prediction of wind pressure coefficients on building surfaces using artificial neural networks. Energy and Buildings, 158, 1429–1441. https://doi.org/10.1016/j.enbuild.2017.11.045

Bronicki, L. Y., 2013. Geothermal Power Stations☆. In Reference Module in Earth Systems and Environmental Sciences. Elsevier. https://doi.org/10.1016/B978-0-12-409548-9.05463-4

Brookshier, P., 2004. Hydropower technology. Encyclopedia of Energy, 3, 333–341.

CAREY, E., 2024. Le Paquet Energie-Climat 2030 de l'Union européenne et la Conférence de Paris sur le climat (COP 21). Mars, 0.

Chakhtoura, Y., 2023. La contribution des énergies renouvelables à l'équilibre du marché de l'électricité en Europe : L'expérience de Vattenfall. Annales des Mines - Responsabilité et environnement, 109(1), 14–18. https://doi.org/10.3917/re1.109.0014

Clements, B., Coady, D., Fabrizio, S., Gupta, S., & Shang, B., 2014. Energy subsidies: How large are they and how can they be reformed? Economics of Energy & Environmental Policy, 3(1), 1–18.

Collard, F., 2015. Les énergies renouvelables. Courrier hebdomadaire du CRISP, 2252–2253(7–8), 5–72. https://doi.org/10.3917/cris.2252.0005

Coulibaly, P., Anctil, F., & Bobée, B., 1999. Prévision hydrologique par réseaux de neurones artificiels: État de l'art. Canadian Journal of Civil Engineering, 26(3), 293–304. https://doi.org/10.1139/l98-069

Global Energy Crisis – Topics. (n.d.). IEA. Retrieved March 30, 2024, from https://www.iea.org/topics/global-energy-crisis

Dastres, R., & Soori, M., 2021a. Artificial Neural Network Systems. International Journal of Imaging and Robotics (IJIR), 21(2), 13–25.

Dastres, R., & Soori, M., 2021b. Artificial Neural Network Systems. International Journal of Imaging and Robotics (IJIR), 21(2), 13–25.

Demraoui, F. Z., Amhamdi, L., & Harrouz, A. / promoteur., 2023. Analyse et efficacité des projets d'énergies renouvelables: Photovoltaïque et éolienne avec RET Screen [Thesis, UNIVERSITE AHMED DRAIA- ADRAR]. https://dspace.univ-adrar.edu.dz/jspui/handle/123456789/8058

Diop, M., 2015. La COP21: Une chance à saisir pour l'Afrique. Géoéconomie, 77(5), 63–71. https://doi.org/10.3917/geoec.077.0063

Farghali, M., Osman, A. I., Mohamed, I. M. A., Chen, Z., Chen, L., Ihara, I., Yap, P.-S., & Rooney, D. W., 2023. Strategies to save energy in the context of the energy crisis: A review. Environmental Chemistry Letters, 21(4), 2003–2039. https://doi.org/10.1007/s10311-023-01591-5

Ferkous, K., 2009. Etude d'une chaine de conversion d'énergie éolienne. Mémoire de Magister, Université Mentouri Constantine, 18. https://www.researchgate.net/profile/Khaled-Ferkous-2/publication/355697310_ETUDE_D'UNE_CHAINE_DE_CONVERSION_D'ENERGIE_EOLIENNE/links/617a036aeef53e51e1f7297c/ETUDE-DUNE-CHAINE-DE-CONVERSION-DENERGIE-EOLIENNE.pdf

Finon, D., 2024. Les politiques climatiques face aux subventions aux énergies fossiles. RetrievedOctober26,fromhttps://www.larevuedelenergie.com/wp content/uploads/2018/10/Politiques-climatiques-subventions-energies-fossiles.pdf

Fontecave, M., & Candel, S., 2024. Quelles perspectives énergétiques pour la biomasse? [PhD Thesis, Académie des sciences (France)]. https://academie-sciences.hal.science/hal-04530487/

Gao, S., Xiong, X., Zhou, Y., & Zhang, J., 2021. Bearing Remaining Useful Life Prediction Based on a Scaled Health Indicator and a LSTM Model with Attention Mechanism. Machines, 9(10), Article 10. https://doi.org/10.3390/machines9100238

Ghiassi, M., Saidane, H., & Zimbra, D. K., 2005. A dynamic artificial neural network model for

forecasting time series events. International Journal of Forecasting, 21(2), 341–362. https://doi.org/10.1016/j.ijforecast.2004.10.008

IEA, 2023. Global Energy Crisis – Topics. (n.d.). Retrieved November 20, from https://www.iea.org/topics/global-energy-crisis

Anonymous 1: https://www.statista.com/statistics/1031138/wind-energy-production-globally/

Grand, D., & Fontecave, M., 2023. Le foisonnement éolien: Les limites d'un mix électrique à forte proportion d'énergies renouvelables intermittentes. Comptes Rendus. Chimie, 26(G1), 63–75. https://doi.org/10.5802/crchim.232

Guan, Y., Yan, J., Shan, Y., Zhou, Y., Hang, Y., Li, R., Liu, Y., Liu, B., Nie, Q., Bruckner, B., Feng, K., & Hubacek, K., 2023. Burden of the global energy price crisis on households. Nature Energy, 8(3), 304–316. https://doi.org/10.1038/s41560-023-01209-8

Gulliver, J. S., & Arndt, R. E., 1991. Hydropower engineering handbook. McGraw-Hill, Inc. https://conservancy.umn.edu/handle/11299/195476

Gustavsson, L., Nguyen, T., Sathre, R., & Tettey, U. Y. A., 2021. Climate effects of forestry and substitution of concrete buildings and fossil energy. Renewable and Sustainable Energy Reviews, 136, 110435. https://doi.org/10.1016/j.rser.2020.110435

Holland, J. H. (1992). Genetic Algorithms. *Scientific American, 267*(1), 66–73. http://www.jstor.org/stable/24939139

IEA, 2024-Hydropower.(n.d.)..RetrievedAugust7,from https://www.iea.org/energy-system/renewables/hydroelectricity

Johnson G. L., 1985. Wind Energy systems. Citeseer

https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=089f05a848d99e2e811606008d3c02c2c24ac828

Jouzel, J., & Euvé, F., 2016. Apports et limites de la COP21. Études, 4, 7–16. https://doi.org/10.3917/etu.4226.0007

Kalair, A., Abas, N., Saleem, M. S., Kalair, A. R., & Khan, N., 2021. Role of energy storage systems in energy transition from fossil fuels to renewables. Energy Storage, 3(1), e135. https://doi.org/10.1002/est2.135

Khashei, M., & Bijari, M., 2010a. An artificial neural network (p, d, q) model for timeseries forecasting. Expert Systems with Applications, 37(1), 479–489. https://doi.org/10.1016/j.eswa.2009.05.044

Khashei, M., & Bijari, M., 2010b. Un modèle de réseau neuronal artificiel (p , d , q) pour la prévision des séries chronologiques. Expert Systems with Applications, 37(1), 479–489. https://doi.org/10.1016/j.eswa.2009.05.044

Kose, R., Ozgur, M. A., Erbas, O., & Tugcu, A., 2004. The analysis of wind data and wind energy potential in Kutahya, Turkey. Renewable and Sustainable Energy Reviews, 8(3), 277–288.

https://doi.org/10.1016/j.rser.2003.11.003

Kumar, I., Tripathi, B. K., & Singh, A., 2023a. Attention-based LSTM network-assisted time series forecasting models for petroleum production. Engineering Applications of Artificial Intelligence, 123, 106440. https://doi.org/10.1016/j.engappai.2023.106440

Kumar, I., Tripathi, B. K., & Singh, A., 2023b. Attention-based LSTM network-assisted time series forecasting models for petroleum production. Engineering Applications of Artificial Intelligence, 123, 106440. https://doi.org/10.1016/j.engappai.2023.106440

Lamy, C., 2022. Étude de l'utilisation de réseaux de neurones artificiels pour des calculs de haute performance dédiés à la modélisation du transport de sources énergétiques [Phdthesis, Université de Bordeaux]. https://theses.hal.science/tel-03908443

Lim, B., & Zohren, S., 2021. Time-series forecasting with deep learning: A survey. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 379(2194), 20200209. https://doi.org/10.1098/rsta.2020.0209

Lin, J., Ma, J., Zhu, J., & Cui, Y., 2022. Short-term load forecasting based on LSTM networks considering attention mechanism. International Journal of Electrical Power & Energy Systems, 137, 107818. https://doi.org/10.1016/j.ijepes.2021.107818

Liu, B., Zeng, X., & Dong, P., 2021. Malicious URL Detection System Based on LSTM and Attention Mechanism. Journal of Physics: Conference Series, 2037(1), 012016. https://doi.org/10.1088/1742-6596/2037/1/012016

NASA, 2024. Long-Term Warming Trend Continued in 2017, NOAA - NASA. (n.d.). Retrieved March,from https://www.nasa.gov/news-release/long-term-warming-trend-continued-in-2017-nasa-noaa/

Lund, J. W., & Toth, A. N., 2021. Direct utilization of geothermal energy 2020 worldwide review. Geothermics, 90, 101915. https://doi.org/10.1016/j.geothermics.2020.101915

Ma, Q., 2020. Comparison of ARIMA, ANN and LSTM for Stock Price Prediction. E3S Web of Conferences, 218, 01026. https://doi.org/10.1051/e3sconf/202021801026

McKendry, P., 2002. Energy production from biomass (part 1): Overview of biomass. Bioresource Technology, 83(1), 37–46.

Moltke, A. von, McKee, C., Morgan, T., & Töpfer, K. (Eds.)., 2017. Energy Subsidies: Lessons Learned in Assessing their Impact and Designing Policy Reforms. Routledge. https://doi.org/10.4324/9781351283205

Neves, C., Gonzalez, I., Leander, J., & Karoumi, R., 2018. A New Approach to Damage Detection in Bridges Using Machine Learning. 73–84. https://doi.org/10.1007/978-3-319-67443-8_5

Niu, Z., Zhong, G., & Yu, H., 2021. A review on the attention mechanism of deep learning. Neurocomputing, 452, 48–62. https://doi.org/10.1016/j.neucom.2021.03.091

NREAPs and Progress Reports DATA PORTAL. (n.d.).

OECD, 2022. World Energy Investment 2022. Organisation for Economic Co-operation and

Development.https://www.oecd-ilibrary.org/energy/world-energy-investment-2022_7a0d89ff-en

Osman, A. I., Chen, L., Yang, M., Msigwa, G., Farghali, M., Fawzy, S., Rooney, D. W., & Yap, P.-S., 2023. Cost, environmental impact, and resilience of renewable energy under a changing climate: A review. Environmental Chemistry Letters, 21(2), 741–764. https://doi.org/10.1007/s10311-022-01532-8

Qiu, J., Wang, B., & Zhou, C., 2020. Forecasting stock prices with long-short term memory neural network based on attention mechanism. PLOS ONE, 15(1), e0227222. https://doi.org/10.1371/journal.pone.0227222

Rangapuram, S. S., Seeger, M. W., Gasthaus, J., Stella, L., Wang, Y., & Januschowski, T., 2018. Deep state space models for time series forecasting. Advances in Neural InformationProcessingSystems,31. https://proceedings.neurips.cc/paper/2018/hash/5cf68969fb67aa6082363a6d4e6468e2-Abstract.html

Romei, V., & Smith, A., 2023. Global inflation tracker: See how your country compares on rising prices. Financial Times. https://www.ft.com/content/088d3368-bb8b-4ff3-9df7-a7680d4d81b2

Roos, A., 2024. Renewing the Subterranean Energy Regime? How Petroculture Obscures the Materiality of Deep Geothermal Energy Technology in Sweden. Ecological Economics, 219, 108129. https://doi.org/10.1016/j.ecolecon.2024.108129

Sang, S., & Li, L., 2024. A Novel Variant of LSTM Stock Prediction Method Incorporating Attention Mechanism. Mathematics, 12(7), Article 7. https://doi.org/10.3390/math12070945

Sebastian, A., & Tantia, Dr. V., 2025. Multi-variate LSTM with attention mechanism for the Indian stock market. International Journal of Information Management Data Insights, 5(2), 100350. https://doi.org/10.1016/j.jjimei.2025.100350

Sen, R., Yu, H.-F., & Dhillon, I. S., 2019. Think globally, act locally: A deep neural network approach to high-dimensional time series forecasting. Advances in Neural Information Processing Systems,32. https://proceedings.neurips.cc/paper/2019/hash/3a0844cee4fcf57de0c71e9ad3035478-Abstract.html

Shams, S. R., Jahani, A., Kalantary, S., Moeinaddini, M., & Khorasani, N., 2021. The evaluation on artificial neural networks (ANN) and multiple linear regressions (MLR) models for predicting SO2 concentration. Urban Climate, 37, 100837. https://doi.org/10.1016/j.uclim.2021.100837

Siami-Namini, S., Tavakoli, N., & Siami Namin, A., 2018. A Comparison of ARIMA and LSTM in Forecasting Time Series. 2018 17th IEEE International Conference on Machine

Learning and Applications (ICMLA), 1394–1401. https://doi.org/10.1109/ICMLA.2018.00227

Smyl, S., 2020. A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. International Journal of Forecasting, 36(1), 75–85. https://doi.org/10.1016/j.ijforecast.2019.03.017

Tai, B., Yu, L., Huang, Y., Wang, J., Wang, Y., Zhu, Y., Duan, S., Liu, S., Chen, Q., & Gu, M., 2023. Power prediction of photovoltaic power generation based on LSTM model with additive Attention mechanism. 2023 7th International Conference on Smart Grid and Smart Cities (ICSGSC), 474–480. https://doi.org/10.1109/ICSGSC59580.2023.10319231

Tarmanini, C., Sarma, N., Gezegin, C., & Ozgonenel, O., 2023. Short term load forecasting based on ARIMA and ANN approaches. Energy Reports, 9, 550–557. https://doi.org/10.1016/j.egyr.2023.01.060

Tétart, F., 2023. De l'urgence d'une transition énergétique. Atlas, 92–93.

Torres, J. F., Hadjout, D., Sebaa, A., Martínez-Álvarez, F., & Troncoso, A., 2021. Deep Learning for Time Series Forecasting: A Survey. Big Data, 9(1), 3–21. https://doi.org/10.1089/big.2020.0159

Unies, N., 2020. Objectifs de développement durable. New York: Nations Unies. https://dictionary.tn/amp/objectifs-de-developpement-durable/

Viñuales, J. E., 2023. La guerre en Ukraine et la transition énergétique. RED, 5(1), 123–129. https://doi.org/10.3917/red.005.0123

Voyant, C., 2011. Prédiction de séries temporelles de rayonnement solaire global et de production d'énergie photovoltaïque à partir de réseaux de neurones artificiels [Phdthesis, Université Pascal Paoli]. https://theses.hal.science/tel-00635298

Wagner, H.-J., 2020. Introduction to wind energy systems. EPJ Web of Conferences, 246, 00004. https://www.epj-conferences.org/articles/epjconf/abs/2020/22/epjconf_lnes2020_00004/epjconf_lnes2020_00004.html

Wang, S., Li, C., & Lim, A., 2021. Why Are the ARIMA and SARIMA not Sufficient (arXiv:1904.07632). arXiv. https://doi.org/10.48550/arXiv.1904.07632

Wang, W., Yuan, B., Sun, Q., & Wennersten, R., 2022. Application of energy storage in integrated energy systems—A solution to fluctuation and uncertainty of renewable energy. Journal of Energy Storage, 52, 104812. https://doi.org/10.1016/j.est.2022.104812

Wang, Z., & Wang, Z., 2015. La contribution nationale de la Chine en vue de la COP21: État des lieux des engagements chinois à l'horizon 2030. Techniques Financières et Développement, 120(3), 15–18. https://doi.org/10.3917/tfd.120.0015

Anonymous 2 : https://wmo.int/news/media-centre/wmo-confirms-2024-warmest-year-record-about-155degc-above-pre-industrial-level

Yasmeen, R., & Shah, W. U. H., 2024. Energy uncertainty, geopolitical conflict, and militarization matters for Renewable and non-renewable energy development: Perspectives from G7 economies. Energy, 306, 132480. https://doi.org/10.1016/j.energy.2024.132480

Yu, Y., Si, X., Hu, C., & Zhang, J., 2019. A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. Neural Computation, 31(7), 1235–1270. https://doi.org/10.1162/neco_a_01199

Zhang, G. P., 2007. A neural network ensemble method with jittered training data for time series forecasting. Information Sciences, 177(23), 5329–5346. https://doi.org/10.1016/j.ins.2007.06.015

Zhang, Y., 2019. ATTAIN: Attention-based Time-Aware LSTM Networks for Disease Progression Modeling. In Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI-2019), Pp. 4369-4375, Macao, China. https://par.nsf.gov/biblio/10136491-attain-attention-based-time-aware-lstm-networks-disease-progression-modeling

Zhao, D., Zhang, H., Cao, Q., Wang, Z., He, S., Zhou, M., & Zhang, R., 2022. The research of ARIMA, GM(1,1), and LSTM models for prediction of TB cases in China. PLOS ONE, 17(2), e0262734. https://doi.org/10.1371/journal.pone.0262734

Zhou, F., Li, L., Zhang, K., Trajcevski, G., Yao, F., Huang, Y., Zhong, T., Wang, J., & Liu, Q., 2020. Forecasting the Evolution of Hydropower Generation. Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2861–2870. https://doi.org/10.1145/3394486.3403337

CURRICULUM VITAE

I'm Abdel Hack BIO BOULOU, a Master's degree student in Statistics from Çukurova University (Türkiye) and a holder bachelor's degree in applied Statistics from Parakou University (Benin). I have a strong academic background complemented by professional experience in data analysis and stock management, notably at SODECO-SA, AIC, and the non-profit organization AXE-Analysis. I also completed an internship at the Departmental Directorate of Planning and Development of Borgou. In addition, I supported the French rescue team Rapid-France as an interpreter and rescue worker during the 6th February earthquake in Türkiye. Skilled in Python, time series forecasting, predictive modeling, and machine learning, I am proficient with tools such as PowerBI, Stata, Python, and Google Colab. In terms of languages, I'm fluent in French, has professional proficiency in English, and an intermediate level in Turkish.