الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي والبحث العلمي

BADJI MOKHTAR ANNABA-UNIVERSITY
Faculty of Technology
Department of Computer Science

جامعة باجي مختار – عنابة
كلية التكنولوجيا
قسم الاعلام الآلي

Domain: Mathematics and Computer Science
Branch: Computer Science
Specialty: Embedded Computing

Towards a Smart and Efficient Weather Forecasting Model Based on Deep Learning

A Dissertation Submitted to the Department of Computer Science in Partial Fulfillment of the Requirements for the LMD Doctorate Degree in Computer Science.

Submitted by
Zouaidia Khouloud

Supervised by
Pr. Ghanemi Salim

Co-Supervised by
Dr. Bougueroua Lamine

Board of Examiners

Quality	Name	University Degree	Affilitation
Chairman	Boudour Rachid	Professor	Badji Mokhtar Annaba
Supervisor	Ghanemi Salim	Professor	Badji Mokhtar Annaba
Co-Supervisor	Bougueroua Lamine	Professor	Ecole EFREI-Paris
Examiner	Khetatba Mourad	MCA	Badji Mokhtar Annaba
Examiner	Benmohammed Mohamed	Professor	Constantine 2, Constantine

Academic Year: 2022/2023

Abstract

The weather is one of the most crucial components that surround and impact almost everything in our everyday lives. Several studies on this subject have been conducted in order to determine practical strategies to exploit the power of nature for our benefit. or, at least, to reduce its tragic disasters. This is what gave origin to the modern science we know today under the name of weather forecasting (WF). Several global climate affecting features are combined to impact the weather, and the most challenging one is the wind parameter with its flactuated nature and unpredictable behaviour.

Due to nowadays energies crisis, the world will witness a remarkable progress in the alternative energies field, where the wind energy industry is considered as the fastest growing renewable energy, due to its widespread availability and affordable cost, and it is expected to grow even faster in the few coming years, being the most lucrative form of green energy sources.

The aim of our research is to implement a reliable model based on deep learning techniques, capable of handling the complex wind characteristics, and also the investigation of sites suitability for wind farms hosting projects.

Keywords: Weather forecasting ,wind speed forecasting, Energy Consumption, Deep learning, Artificial intelligence

ملخص

يعد الطقس أحد أهم العوامل التي تحيط بنا وتؤثر تقريبًا على كل الجوانب في حياتنا اليومية. تم إجراء دراسات مختلفة في هذا الموضوع من أجل إيجاد طريقة لإسغلال قوة الطبيعة في صالحنا أو على الأقل ، لمنعها من التسبب في الكوارث الوخيمة التي تودي بحياتنا و ممتلكاتنا .

من هنا تمخض العلم المعروف اليوم باسم علم التنبؤ بالطقس حيث تجتمع العديد من العوامل الطبيعية للتأثير عليه ، وأكثرها تعقيدًا هو سرعة الرياح و ذلك راجع إلى طبيعة الرياح المتقلبة وسلوكها الغير المتوقع.

بسبب أزمة الطاقة في الوقت الحاضر ، يشهد العالم تقدمًا ملحوظًا في مجال الطاقات المتجددة ، حيث يُصنف توليد طاقة الرياح كواحد من الطاقات البديلة الأسرع نموًا ، ومن المتوقع أن يشهد انتشارا أسرع في السنوات القادمة ، حيث يعد من مصادر الطاقة الخضراء الأكثر ربحية بسبب توفره بشكل مستمر على مدار السنة وتكلفته المنخفضة.

الهدف من بحثنا هذا هو تطوير نموذج فعال و موثوق يعتمد على تقنيات التعلم العميق ، وقادر على التعامل مع خصائص الرياح المعقدة ، وكذلك التحقيق في ملاءمة المواقع المقترحة لاستضافة المشاريع المستقبلية لبناء مزارع لتوليد طاقة الرياح.

الكلمات المفتاحية: التنبؤ بالطقس ، التنبؤ بسرعة الرياح، استهلاك الطاقة ، التعلم العميق ، الذكاء الاصطناعي،

Résumé

La météo est l'un des éléments cruciaux qui entourent et influencent presque tout dans notre vie quotidienne. Différentes études ont été menées sur ce sujet afin de trouver un moyen de bénéficier du pouvoir de la nature ou, du moins, de l'empêcher de provoquer des catastrophes.

C'est ce qui a donné naissance à la science connue sous le nom de prévision météorologique (PM). De nombreux facteurs d'influence du climat mondial se combinent pour influencer l'atmosphere, et le plus complexe est la vitesse du vent à cause de sa nature fluctuante et son comportement imprévisible.

En raison de la crise énergétique actuelle, le monde va connaître un progrès remarquable dans le domaine des énergies renouvelables, où la production d'énergie éolienne est classée comme l'une des énergies alternatives à la croissance la plus rapide, et elle devrait croître encore plus rapidement dans les années à venir, étant la source d'énergie verte la plus rentable en raison de sa disponibilité et de son coût raisonable.

L'objectif de notre recherche est de développer un modèle fiable basé sur des techniques d'apprentissage profond, capable de gérer les caractéristiques complexes du vent, ainsi que l'étude de l'adéquation des sites pour les projets d'hébergement de parcs éoliens.

Mots-clés: Prévision météorologique, prévision de la vitesse du vent, consommation énergétique , apprentissage profond, intelligence artificielle

Acknowledgements

First, I would like to thank Allah, for awarding me with his immeasurable blessings at every stage of my journey toward the completion of my research work successfully .

I would like to express my heartfelt appreciation to my advisor Prof. Ghanemi Salim, for his immense knowledge that he never skimp me from, for his patience, motivation, his permanent availability , for his guidance over the past three years and for shaping me to grow as a researcher.

Working under your supervision was an honor.

I am also in debt to my co-advisor Dr. Bougueroua Lamine and to Dr. Feriel Bouakkaz who gave me the opportunity to award a scholarship at the Efrei Paris-Pantheon-Assas University Paris II.

With the help of this scholarship, I was able to complete my thesis, carry out my research, meet experts from different research fields, and spend four months in Paris to return home with a lot of wonderful memories that I will treasure forever.

I appreciate as well the efforts made by the reviewers and the time dedicated to read and review this thesis.

I would like to express my gratitude to mom and dad for their unconditional encouragement and support throughout this research journey. Thank you for being such great parents. I am so lucky to have you both

Special thanks to my sisters "Manel", "Nour" and little "Yara" , the only ones I ever trust and think of to share with the good and specially the bad, I am blessed with you my besties. Thank you for lifting me up when I couldn't stand on my own.

Finally, to my partner in crime, without you this work would never seen the light, thank you for putting your trust on me, for your support and for seeing in me what no one else could see, am so blessed with you and so grateful for the great memories we have shared during this Phd journey.

Contents

List of Figures

List of Tables

Abbreviations

ADMM	**A**lternate **D** irection **M**ethod **M**ultipliers
AVMD	**A**daptive **V**ariational **M**ode **D**ecomposition
Bi	**Bi**directional
CEEMD	**C**omplete **E**nsemble **E**mpirical **M**ode **D**ecomposition
CF	**C**entral **F**requency
CMA	**C**entered **M**oving**A**verage
CNN	**C**onvolutional **N**eural **N**etwork
CSO	**C**uckoo **S**earch **O**ptimization
CWT	**C**ontinuous **W**avelet **T**ransform
DBSCAN	**D**ensity **B**ased **S**patial **C**lusteringl of **A**pplications with **N**oise
DFF	**D**eep **F**eed **F**orward
DFT	**D**iscrete **F**ourier **T**ransform
DirREC	**Dir**ect **Rec**ursive
DNN	**D**eep **N**eural **N**etwork
DWT	**D**iscrete **W**avelet **T**ransform
ED	**E**ncoder **D**ecoder
EEMD	**E**nsemble **E**mpirical **M**ode **D**ecomposition
EFG	**E**nhanced **F**orget gate **N**etwork
ELM	**E**xtreme **L**earning **M**achine
EMA	**E**xponential **M**oving **A**verage
EMD	**E**mpirical **M**ode **D**ecomposition
FEEMD	**F**ast **E**nsemble **E**mpirical **M**ode **D**ecomposition
FT	**F**ourier **S**eries

FFT	**F**ast **F**ourier **T**ransform
FNN	**F**eed Foreward **N**eural **N**etwork
FWT	**F**ast **W**avelet **T**ransform
GRU	**G**ated **R**ecurrent **U**nit
GW	**G**ega **W**att
GWO	**G**rey **W**olve **O**ptimizer
kmph	**k**ilometer **p**er **h**our
LSTM	**L**ong **S**hort **T**erm **M**emory
MA	**M**oving **A**verage
MAE	**M**ean **A**bsolute **E**rror
MAPE	**M**ean **A**bsolute **P**ercentage **E**rror
mb	**m**illi **b**ars
MEMD	**M**ultivariate **E**mpirical **M**ode **D**ecomposition
MIMO	**M**ulti **I**nput **M**ulti **O**utput
MISMO	**M**ulti **I**nput **S**everal **M**ulti **O**utput
MLP	**M**ulti **L**ayer **P**erceptron
mph	**m**iles **p**er **h**our
MSE	**M**ean **S**quared **E**rror
MW	**M**ega **W**att **h**our
NWP	**N**umerical **W**eather **P**rediction
OVMD	**O**ptimal **V**ariational **M**ode **D**ecomposition
RBFNN	**R**adial **B**asis **F**unction **N**eural **N**etwork
REI	**Re**sidual **I**ndex
ReLu	**Re**ctified **L**inear **u**nits
RMSE	**R**oot **M**ean **S**quare **E**rror
RMSprop	**R**oot **M**ean **S**quared **prop**propagation
RNN	**R**ecurrent **N**eural **N**etwork
SDAE	**S**tacked **D**enoising **A**uto **E**ncoders
SMA	**S**imple **M**oving **A**verage
SNN	**S**hallow **N**eural **N**etwork
SSA	**S**ingular **S**pectrum **A**nalysis

SVD	**S**ingular **V**alue **D**ecomposition
SVR	**S**upport **V**ector **R**egression
TMA	**T**railing **M**oving **A**verage
TVF-EMD	**T**ime **V**arying **F**ilter-based **E**mpirical **M**ode **D**ecomposition
TWh	**T**era **W**att **h**our
VMD	**V**ariational **M**ode **D**ecomposition
WMA	**W**eighted **M**oving **A**verage
WT	**W**avelet **T**ransform

To

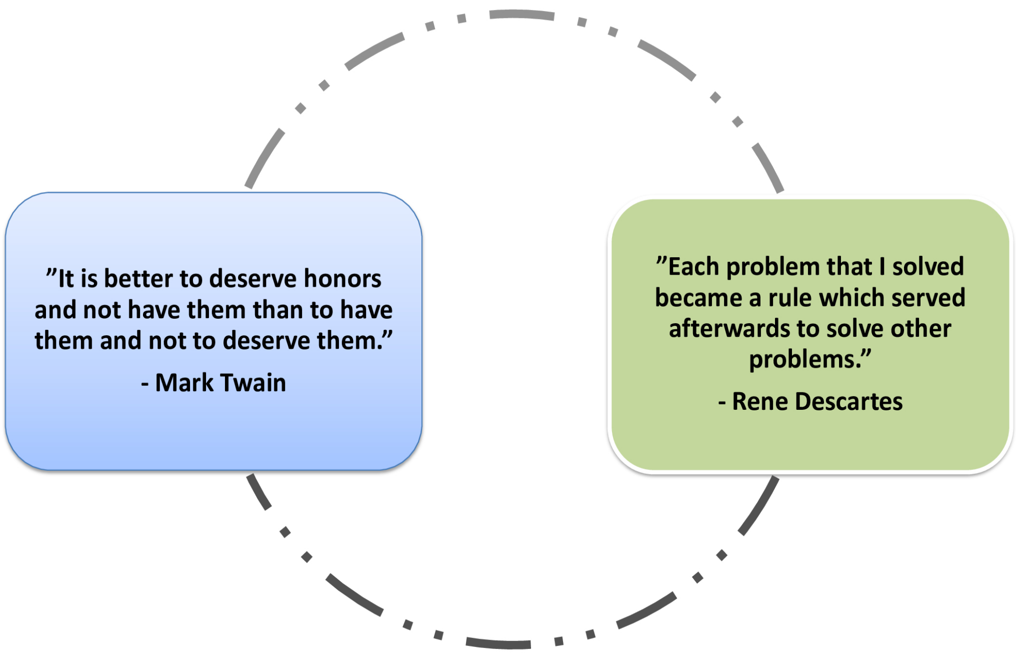
My dear parents . . . My lovely sisters

My partner in crime

and all my family and friends in honor of their value, love, and efforts that helped to shape who I am today. . . .

Chapter 1

General Introduction



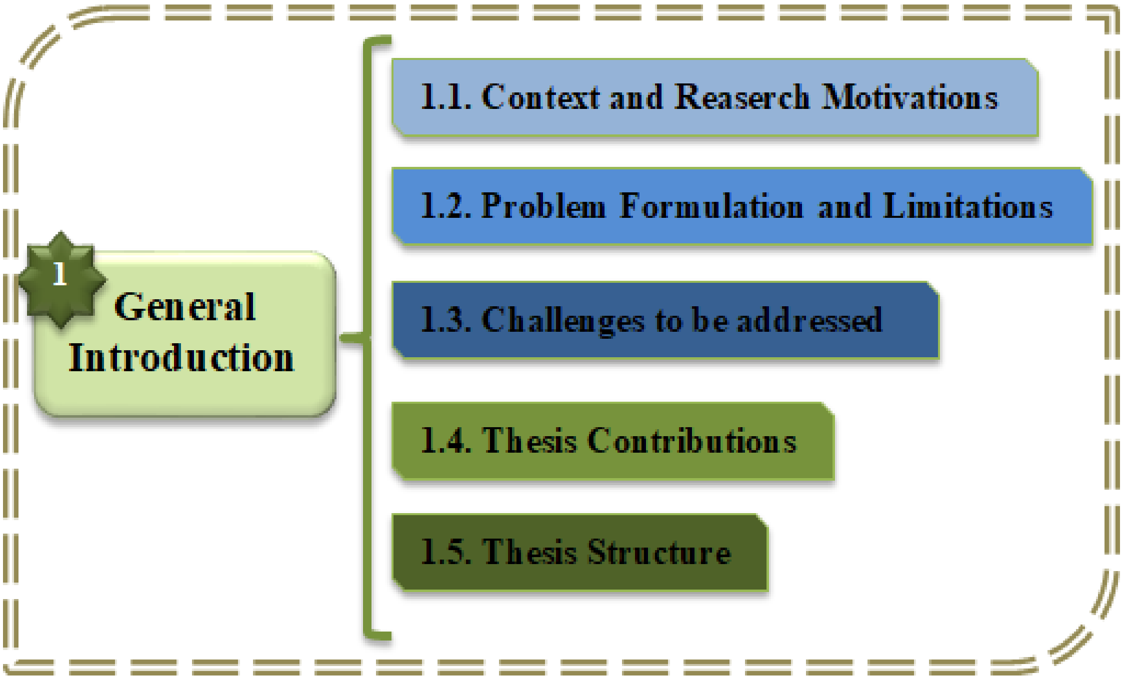


FIGURE 1.1: Chapter1 Map.

Nowadays energy consumption significantly increased, as a side effect of the expansion of manufacturing facilities, and the rapid growth of economic activities.

Since wind power is one of the most economical sources of electricity, it is regarded as the ideal response to the significant increase in energy demand. Therefore, in the upcoming years, the number of hosted wind farms must increase [1].

Recent studies have revealed that the key to establish reliable power generating plans is to accurately estimate the wind speed feature.

Numerous methods for predicting wind speed have recently been proposed, including: physical techniques that use numerical weather prediction (NWP) to prognosticate medium and long-term forecasting times scales[2], and deep learning techniques for short and ultra-short time horizons that proved their effectiveness by outperforming statistical traditional models that typically use recursive linear models.

This thesis investigates the use of hybrid deep learning approaches to solve problems and find alternatives to weather forecasting tasks, specifically wind speed forecasting, where the objectives of our study were to understand the forecasting mechanism, and to develop an intelligent framework that handles the dependency of the existing models on different data characteristics which is consuming in both time and calculation.

1.1 Context and Motivation

Climate changes addresses one of our time's most critical social and economic concerns, and it is expected to increase the frequency and intensity of numerous extreme weather disasters. Citizens are actively encouraging their governments to handle one of today's major challenges by allowing investments in wind power generation, the most cost-effective climate mitigation technology.

The increased growth of manufacturing and the acceleration of commercial development resulted in a significant rise in energy consumption. While the world continues to suffer from environmental degradation and global warming caused by the mismanagement of conventional energy supplies such as coal and natural gas. As a result, Today, more than ever, the world requires alternative energy supplies, and wind power generation is regarded among the most profitable sustainable energies in terms of materials and costs.

Wind energy capacity increased by 17% last year, and it is currently the leading source of alternative energy in the most advanced nations.

Algeria's government has launched a captivating strategy to attract investment in wind power, with a target of establishing 5 GW of wind energy by 2030.

1.2 Problem Formulation and Limitations

Due to its complex nature and difficult-to-predict parameters, the wind is one of the most challenging weather events to forecast.

We chose the western regions of Algeria, characterized by the highest records of wind speed in the entire nation, showing a significant potential for the proposed research, with evaluating the proposed models on multiple prominent datasets with various features.

For the ideal hosting strategy of future wind farms, in this thesis we will concentrate on developing an effective wind speed forecasting system, in order to record the greatest and most steady wind speed values in Algeria.

For the first time, the issue of the preprocessing approaches dependency on the various dataset features was addressed.

The main limitation of our investigation was the lack of tools for validation and real-time verification of the developed models on real wind farms systems.

1.3 Challenges to be addressed

Forecasting wind parameters remains a difficult task due to the wind's extremely unpredictable nature, that necessitates a reliable framework that has the ability to automatically learn characteristics from a sequence based on its temporal ordering. Recent studies have shown that predicting the wind speed is crucial key in building reliable power generation frameworks, where deep learning strategies outperformed traditional models.

The primary challenges treated in this thesis are :

- The determination of the ideal combination for the suggested framework to take advantage of the significant time series characteristics, improve forecasting precision, and shorten the computation time

- To build a common architecture that can handle the variations in the datasets.

- To achieve the best forecasting optimization, combining the preprocessing techniques that heavily depend on the dataset characteristics of each research area in an effective way with the structure of the prediction model.

1.4 Contributions

The primary contributions of this thesis are outlined in this section, where we attempted to :

• Find the most appropriate combinations for the proposed hybrid architectures in order to take advantage of the time series key features and improve the prediction performance.

• Create a standard framework that can handle multiple time series without adaptation, to overcome the restrictions of the hyper-parameter selection process, that is consuming in computation and time.

• Select the most prominent zones that grant wind speed stability and predictability for wind farms hosting projects .

1.5 Publications

1.5.1 Journal Publications

1. **Zouaidia, K.**,et al. (2021). Hybrid intelligent framework for one-day ahead wind speed forecasting. Neural Computing and Applications, 33(23), 16591-16608.

Classe: A, Impact Factor: 5.606, H-index: 96.

2. **Zouaidia, K.**,et al. (2023). Weather forecasting based on hybrid decomposition methods and adaptive deep learning strategy. Neural Computing and Applications, 1-16.

Classe: A, Impact Factor: 5.606, H-index: 96.

3. **Zouaidia, K.**, et al. (2021). Hourly Wind Speed Forecasting Using FFT-Encoder-Decoder-LSTM in South West of Algeria (Adrar). International Journal of Informatics and Applied Mathematics, 4(1), 72-83.

1.5.2 Other Journal Publications in parallel with the thesis

1. Rais, M. S., **Zouaidia, K.**, and Boudour, R. (2022). Enhanced decision making in multi-scenarios for autonomous vehicles using alternative bidirectional Q network. Neural Computing and Applications, 1-16.

Classe: A, Impact Factor: 5.606, H-index: 96.

2. Rais, M. S., Boudour, R., **Zouaidia, K.**, and Bougueroua, L. (2022). Decision making for autonomous vehicles in highway scenarios using Harmonic SK Deep SARSA. Applied Intelligence, 1-18.

Classe: A, Impact Factor: 5.019, H-index: 72.

1.5.3 International Conferences Publications

1. **Zouaidia, K.**,et al. (2020, December). Wind speed forecasting based on discrete wavelet transform, moving average method and gated recurrent Unit. In International Conference in Artificial Intelligence in Renewable Energetic Systems (pp. 71-78). Springer, Cham.

2. **Khouloud, Z.**, et al. (2021, September). Multi-Step Wind Speed Forecasting Based on Hybrid Deep Learning Model and Trailing Moving Average Denoising Technique. In 2021 International Conference on Recent Advances in Mathematics and Informatics (ICRAMI) (pp. 1-5). IEEE.

1.5.4 National Conferences Publications

1. **Zouaidia, K.**, GHANEMİ, S., and RAİS, M. S. (2020) Hourly Wind Speed Forecasting using FFT-Encoder-Decoder-LSTM in South West of Algeria (Adrar) in Third Conference on Informatics and Applied Mathematics, IAM's 2020.

2. Rais M.S., Boudour, R., and **Zouaidia, K.**. (2020).Avoiding obstacles in a road with a maze structure using reinforcement learning methods in Third conference on informatics and applied mathematics, IAM's 2020.

1.6 Thesis Structure

This dissertation is divided into four chapters, besides the chapter of the general introduction. The second and third chapters describe the theoretical background and foundations of the approaches proposed in this thesis while the fourth chapter is dedicated to describe the primary contributions of this thesis. The last chapter provides the general conclusion, and anticipates future perspectives.

In the first chapter we have developed the motivations for this research and its context followed by the problem formulation, the limitations and the challenges addressed by this work.

Then, we summarised the main contributions from this research, and finally, we described the structure of the thesis.

The second chapter (Background), introduces the three research axis that inspired our research topic, named weather forecasting, Deep Learning and renewable energies. It discusses the essential points concerning these concepts, starting with the necessary fundamentals to familiarize with the different terms of the three axis, reaching to a synthesis that was the main stone for the choices of the upcoming publications.

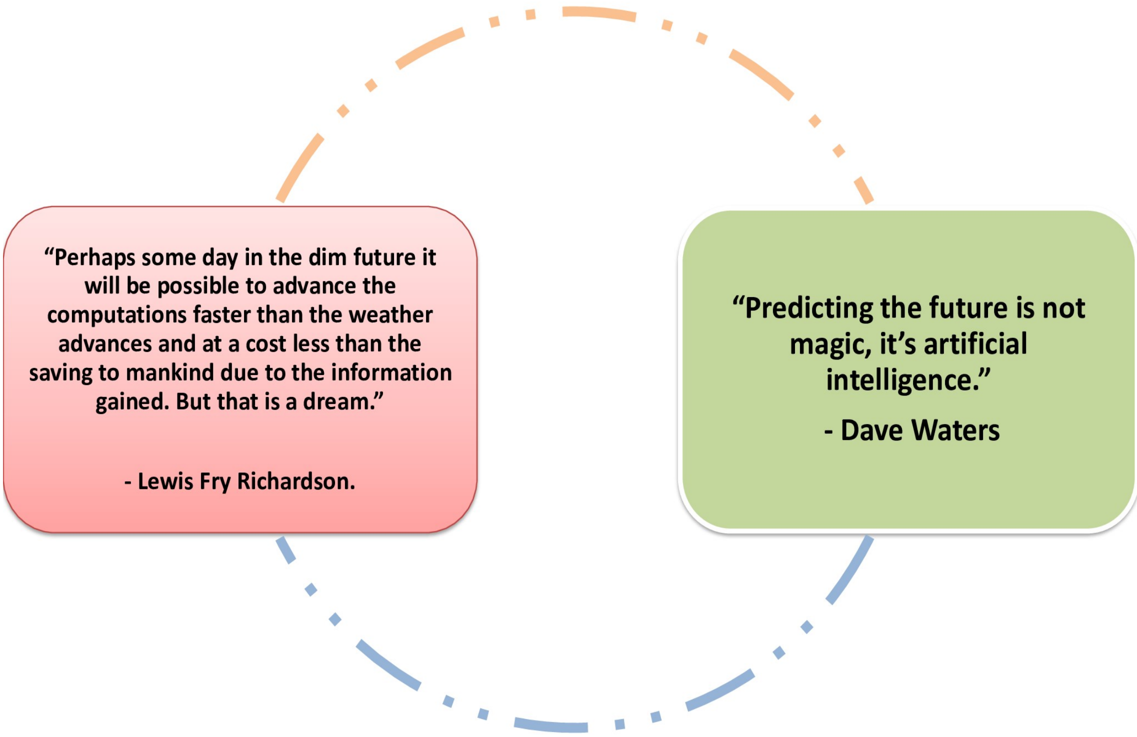
In the third chapter (Wind Speed Forecasting), we further explore a selection of the most popular data preprocessing techniques, relevent optimization algorithms, and deep learning models used in the wind speed forecasting area, along with some pertinent researches.

The fourth chapter (Contributions), describes our achievements and contributions during the thesis investigation.

The conclusion emphasizes the benefits of the proposed approaches, while also discusses the results obtained throughout this thesis and the perspectives to be assigned to this work.

Chapter 2

Background



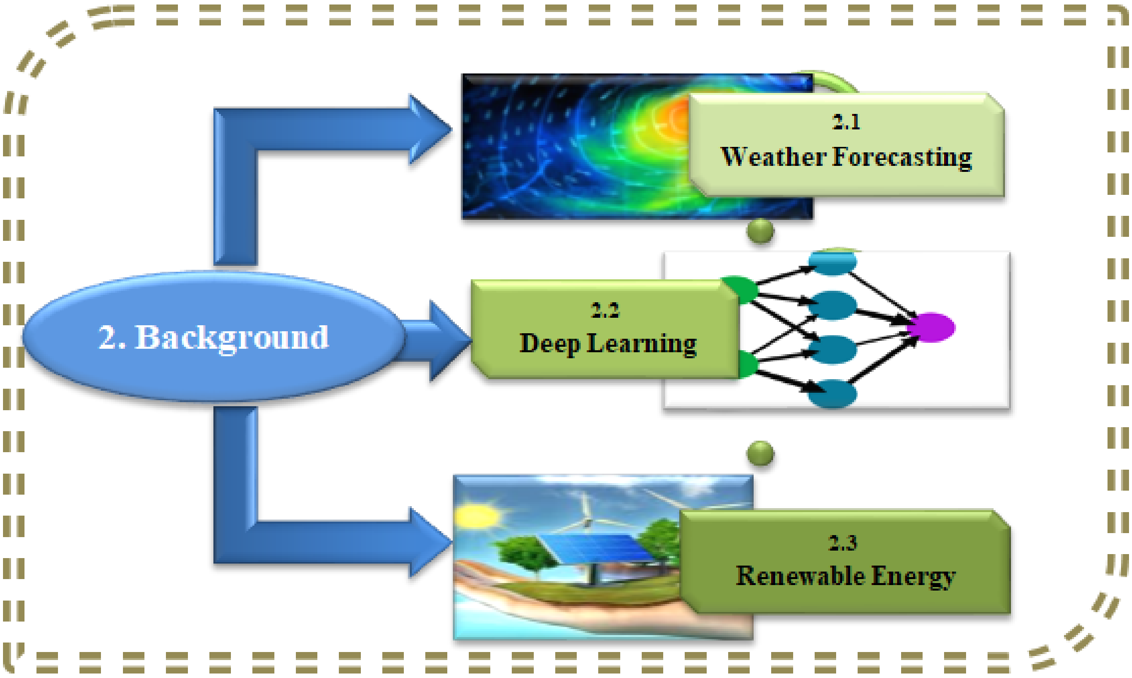


FIGURE 2.1: Chapter2 Map.

2.1 Introduction

The application of advanced technologies and relevant scientific approaches to predict the atmosphere conditions at a given location is known as weather forecasting, created by collecting quantitative data about the different weather parameters and applying scientific theories of atmospheric processes to forecast the future changes possible to occur in the atmosphere [3].

Forecasting in the past is based initially on the observation of weather key features. The study of the different atmospheric parameters has led to the development of a variety of techniques [4], ranging from relatively straightforward sky observation to extremely sophisticated computerized mathematical models.

Thanks to the emergence of deep learning methods, the wide availability of massive weather observation information, as well as the advent of data and sophisticated technology, many researchers have been motivated to investigate hidden hierarchical features in considerable volumes of meteorological information for weather future prediction, using sophisticated hybrid deep learning architectures.

When applied to renewable energy sources, weather forecasting becomes even more vital.

The potential impact of weather on supply capabilities will increase as the world shifts to weather-dependent energy sources such as wind power generation.

Today, weather complex conditions are causing significant fluctuations in energy production.

In this context, efficient weather forecasting process has a key role to play in the energy transition and green electricity grids .

This chapter covers the theory and fundamentals behind the research axis that inspired our research topic. It provides the evolution of weather forecasting approaches and its intersection with deep learning techniques and renewable energies.

2.2 Weather Forecasting :

Current and upcoming weather analysis have always been a great interest, in the management of safe air and ship traffic, in agriculture and forestry, in the effective production of energy using sustainable technologies, as well as in transportation systems. All depend on accurate weather forecasts.

The use of advanced technologies and sophisticated materials to forecast the the future meteorological conditions for a specific location and duration is known as weather forecasting. This covers the temperature, wind speed, wind direction, precipitation, and humidity [5].

2.2.1 Meteorological factors

Bellow a brief description of some important weather factors for predicting weather conditions.

Temperature:

The most significant intensive property is temperature, which is also a meteorological parameter that quantifies the measure of hotness or coldness in a specific location at a specific time (or can be taken as the average of the values recorded up to that point), and it is expressed in terms of arbitrary scales. Several tools, including thermometers, Stevenson screens, and thermo-hygrographs, are used to measure temperature [6].

Pressure:

The force exerted by the weight of air on a surface is defined as pressure. It can be measured in millibars (mb) using devices such as an aneroid barometer, a digital barometer, and a barograph [7].

Humidity:

The amount of water vapor in the air is referred to as humidity. The air is so densely packed with water vapor that there is no space for other particles. Typically, it is expressed as a percentage [8].

Precipitation:

Precipitation is the amount of water that falls from the clouds as rain or snow to the ground. It is measured in millimeters using a device known as a rain gauge.

Wind Speed:

Wind speed is described as the basic atmospheric rate where the air moves in the atmosphere. It is considered as the hardest parameter to predict that is why we chose it as the main focus of this thesis. It can be measured using instruments such as an anemometer and a wind vane, and can be represented in kilometer per hour (kmph) or miles per hour (mph) [9].

2.2.2 Traditional Meteorological Prediction Methodologies

In the past, people have made efforts to comprehend how the atmosphere behaves by examining the patterns and connections between phenomena and connecting them to upcoming events.

Weather forecasting has been an ambiguous activity practiced by civilizations since the dawn of human history. The foundation of weather prediction was laid by ancient Greek philosophers, then continued by renaissance scientists, followed by the 17th and 18th century scientific revolution. The theoretical strategies of twentieth- and twenty-first-century with the help of meteorological scientists aided in the advancement of weather forecasting. The synoptic weather map became the primary tool of nineteenth-century meteorologists. This is now used in weather reports on television around the world [10].

Understanding the long-term trend of historical climate records, and applying appropriate approaches will enhance the performance of weather prediction with the least cost, as science has advanced and deep learning has been introduced. Forecasting is typically used to foresee the weather for the upcoming days, that requires a systematic collection of weather records from various locations, using meteorology, data readings, and measurement techniques, with a variety of implementations[11].

Depending on the time factor and the weather component for which the forecast is required, there are various methods used in forecast preparation, and that can be presented as follows:

2.2.2.1 Persistence Weather Prediction

Considered as the simplest method of weather prediction. It uses the conditions of today to predict the weather of tomorrow. When the weather is stable, as it is during the summer, this can be a reliable method of predicting the weather.

There have to be a persistent weather features for this forecasting technique to work. Both short and long-term predictions can profit from it. This presupposes that the current weather will continue to be as it is, where meteorologists make weather observations to learn more about the weather [12].

2.2.2.2 Synoptic Weather Prediction

In this traditional method, a meteorologist regularly creates the synoptic charts by keeping track and compiling information on various weather components from numerous weather stations situated at the proper locations, and operating during the specified time period. Synoptic maps form the basis of weather forecasts and didn't become common until the late 1950s. Numerous methodological rules have been developed over the course of many years as a result of the analysis of synoptic charts, which aid in determining the strength and motion of weather systems [13].

2.2.2.3 Numerical Weather Predictions (NWP)

The numerical weather prediction methodology forecasts weather by using the power of supercomputers to compute complex mathematical and scientific equations that simulate atmospheric conditions [14]. Because of the lack of a proper definition of the initial state, the computer's estimation of how the initial state will evolve is unpredictable.

2.2.2.4 Statistical Methods

The statistical approaches, are mainly focused on the idea that what is coming will be a continuation of the past and thus include weather prediction based on previous weather records, to overcome the limitations of the NWP models.

To foretell the weather in the future, we need to identify reliable predictors of events and create relationships among parameters. It is specifically used to forecast a single weather factor at a time.

The parameters that define weather conditions, such as minimum and maximum wind speed, average rainfall, and so on, change frequently over time, formulating time series of each parameter that can be used to develop a forecasting model that employs this time series data[15].

2.2.3 Time series

A time series is a set of information collected at regular intervals of time.

They are widely used in any domain that requires temporal measurements, and they have a wide range of features. For example, all time-series have a level, most have noise, and only a few have trend and seasonality.

Formally, a time series named X is defined as a discrete sequence of tuples composed of single or multiple observations from one time step.

Each time series has a set interval t between steps, and the steps are named as $t_i \in T$ where $T \in N$.

There are two categories of time series : univariate and multivariate time series. In the first, each observation comprises a single value, while in the following each step is set as a values list[16].

Weather series are generally multivariate because each time step is composed of multiple observations calculated from various sensors.

The following is an illustration of the multivariate weather time series z_i:

$$Z : z_i, z_2, ..., z_n \tag{2.1}$$

Where,

i : list or a tuple of m values.

$$z_i : x_1, x_2, x_3, ..., x_m \tag{2.2}$$

Each list is constituted of multiple observations:

x_1 : wind speed.

x_2 : wind direction.

x_3 : temperature.

x_m : other weather factors.

The time step's observations can be numerical or categorical.

Categorical values are typically used to define qualitative properties and could be totally unconnected.

Numerical values can be either an integer $v \in N$ or a real number $v \in R$.

The different steps in a time series have a sort of connection between each-other, and the frameworks we create aim to determine these relationships [17].

Significant combination can have an obvious relation (a linear relationship), combinations with higher complexity (a non-linear function with multiple values at each step), or possible to have no connections at all when it comes to random mechanism [18].

2.2.3.1 Time series Forecasting

forecasting a time series is a regression problem, similar to the traditional methods of prediction, that consist of developing a model as an empirical representation of the time series, the model will serve as a hypothesis on the probability of the values of the steps to be forecasted [19].

Forecasting is considered as an optimisation problem to find the best model parameters. For this, the model definition is critical and plays a primary role in the prediction performance, and for that it must be experimentally validated to obtain the best fitting approach for each time series characteristic [20].

Various criteria can be used to classify forecasting models. An initial taxonomy of three major classes can be suggested as follows :

- Point forecasting and probabilistic forecasting.

- Linear and non-linear models.

- Single step and multiple step forecasting : that we will further explain in the next section[21].

2.2.3.1.1 Single-step and Multiple-Step ahead forecasting :

Forecasting a single value in the future is referred as single-step forecasting. Generally, the weather forecasting models are single-step, because they are built to achieve single predicted value in the future. While multi-step forecasting is the prediction of multiple steps in the future.

Single-step forecasting is the prediction with horizon H of a single value into the future, as detailed in Equation 2.3.

$$(z_1, z_2, ..., z_t) \rightarrow (\hat{y}_t + H) \tag{2.3}$$

Multi-step ahead forecasting is the process of predicting a time series with horizon H from a time series with $t(1, 2, ..., t)$ past observations [22].

$$(z_1, z_2, .., zt) \rightarrow (\hat{y}_{t+1}\hat{y}_{t+2}, ..., \hat{y}_{t+H}) \tag{2.4}$$

Various methods can be used to generate the multi-step prediction, illustrated in Figure 2.2.

a. Multi-step Recursive (Iterative) Forecast:

Generally viewed as the primary method to study multi-step-ahead forecasting, it involves computing a one-step model multiple times and using the results as an input for the next time step forecast[23].

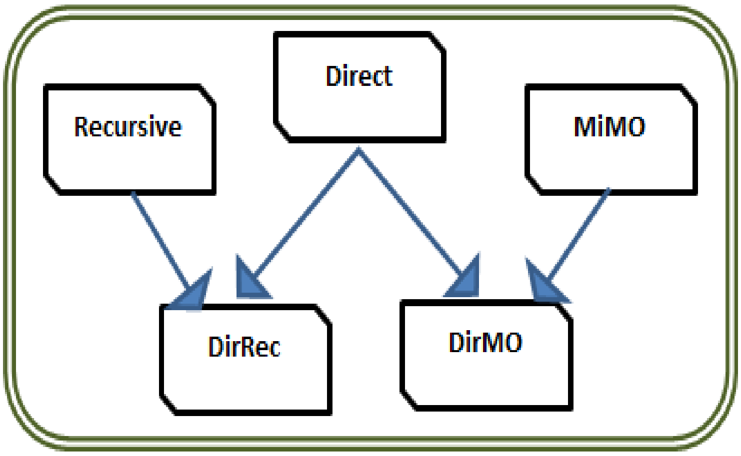


FIGURE 2.2: Multi-Step-ahead prediction.

Additionally, since the model relies on predictions rather than observations, errors can accumulate and worsen the model's performance as the forecasting time range expands.

b. *Direct Multi-Step Forecasting Technique:*

This strategy uses separate models for each forecast time step, which adds computational and maintenance weight, especially when there are many time steps to be forecasted [24].

Another issue to note is that there is no way to model the dependencies between the various predictions.

c. *DirREC :*

This strategy integrates the concepts of both the Direct and Recursive strategies, where it makes use of various models for each time-step of the forecast, and extends the inputs at each step by introducing variables corresponding to the previous step's forecasts.

d. MIMO stands for (Multi-input Multi-output):

The multiple output method predicts the entire forecast sequence using the same model structure, however, it is more complex because it preserves the stochastic dependency of the time series between the estimated parameters. This concept helps to surpass the shortcomings of the preceding strategies but makes the training process slower and requires extra data to avoid overfitting.

e. (MISMO) :Hybrid DirMO :

It is an intermediate strategy that considers two aspects: the dependency of future forecast, while also providing higher predictor flexibility in which the limitation of MIMO approach is eliminated by tuning an integer parameter S, that adjusts the dimensions of the output on the basis of a validation set of criteria [23].

2.3 Deep Learning

Deep learning (DL) models are currently used to address the latest sophisticated challenges, including weather forecasting. As a result, deep learning methods are considered for this research.

First, we would like to review some DL fundamentals.

Artificial neural networks (ANN) are the building blocks of DL (Figure 2.4) , they are known as perceptrons that imitate the same functionality as human brain neuron. Neurons are made up of dendrites, axons, and cell bodies. Dendrites are the receptors of the neurological signal that transfer it then to the cell body, which analyses the inputs and determines whether to send it to other neurons via chemical transmission through axon or not [25].

Artificial neurons, receive (x1,x2,... xm) as input signals,then multiply all inputs by (w1,w2,... wm) as consecutive weight, then sum it using a predetermined bias, and use the activation function f(x) to feed them through.

The signal is produced as 0 or 1 depending on the threshold of the activation function. A single layer perceptron consists of input signals, weights, summation and bias, activation function, and output. In most ANN schemes, only the input and output layers are displayed. Hidden layers are introduced between the input

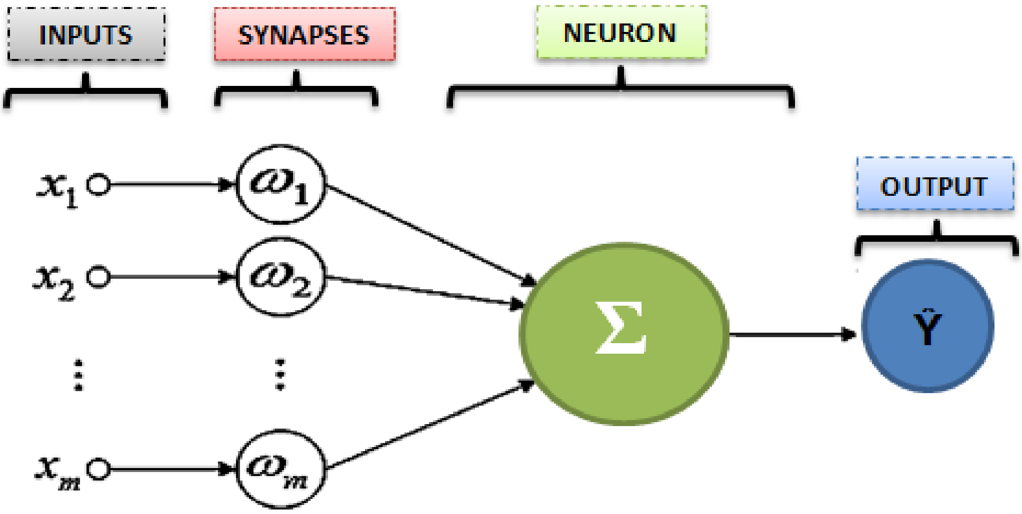


FIGURE 2.3: Simple Artificial Neuron.

and output layers in a real neural network. The hidden layers's number is a hyperparameter that is normally determined by trial and error, and by observing model performance .

A Shallow Neural Network (SNN) is a model with only one hidden layer, whereas a DNN comprises multiple ones. In this study, We studied Convolutional Neural Network (CNN), Encoder Decoder model and Recurrent Neural Network (RNN), in the form of Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) that will be explored more in the upcoming chapter[26].

2.3.1 Deep Neural Network (DNN)

DNN is constructed of three types of layers, an input layer, single or multiple hidden layers, and an output layer. The number of hidden layers is a hyperparameter that must be determined by trial and error [27].

Figure 2.4 represents a DNN architecture with stacked hidden layers, each with four neurons, four inputs, and one output. The number of neurons is proportional to the number of inputs and outputs.

$$f(x; W; c, w, b) = w^T max\{0, W^T + c\} + b \tag{2.5}$$

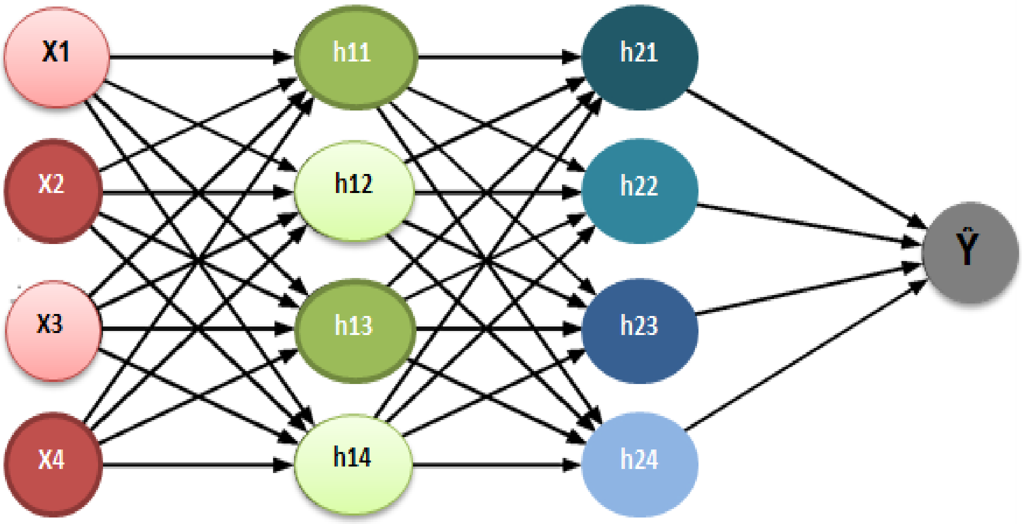


FIGURE 2.4: DNN Model.

Equation 2.5 examines linear modeling and formulates a simpler DNN kernel. Input, weights, and bias are represented by x, W, and c, respectively, whereas w and b are the parameters of the linear model.

Equation 2.6 shows the hidden layer parameter h, where g is the activation function.

$$h = g(W^t x + c) \tag{2.6}$$

ReLu is utilized as the hidden layer activation function in DNN modeling and it is represented in equation 2.7 [28], while further explained in the next chapter.

$$f(x) = max(0, x) \tag{2.7}$$

2.3.2 Time series using deep learning :

Deep learning methods have a significant impact on time series forecasting performance, beginning with the automatic understanding of temporal relationships and progressing to automatically handling temporal patterns such as trends and seasonality [29].

In recent years, many new deep learning models have been investigated in this field in order to increase weather forecasting accuracy and obtain more accurate predictions, as demonstrated by the Recurrent Neural Network (detailed in the next chapter), that proved its efficiency [30], adding native support for input data and handling the order between observations when learning a mapping function from inputs to outputs, and thus outperformed other benchmark models [31].

2.3.2.1 Performance Evaluation criteria of Deep Learning Weather Forecasting Models

Multiple evaluation metrics are chosen to compare the effectiveness of competitive DL models, in order to completely evaluate prediction outcomes in terms of reliability and efficiency.

Prediction accuracy is measured using : Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) .

The usage of MAE and RMSE helps in solving the problem of balancing positive and negative deviations by displaying the average deviation between the anticipated state and the current state.

MAPE is a widely used indicator for the evaluation of the complete forecasting process [32].

The following equations represent the detailed formulation of the evaluation criteria :

$$MAE = (\sum_{t=1}^{N} |X(t) - \hat{X}(t)|)/N \tag{2.8}$$

$$MAPE = (\sum_{t=1}^{N} |(X(t) - \hat{X}(t))/X(t)|)/N \tag{2.9}$$

$$RMSE = \sqrt{(\sum_{t=1}^{N}[X(t) - \hat{X}(t)]^2)/(N-1)} \tag{2.10}$$

Having,
X(t) : as the weather values.
$\hat{X}(t)$: as the forecasted values.
N : as the number of the X(t).

2.3.2.2 Suitability of the Weather Forecasting Models

Because of climate unpredictability, and the challenging behaviour of the wind speed variable, a simple model cannot adequately capture those complicated changes. As a result, the chosen forecasting model should be capable of modeling the complicated nonlinear connection between the input characteristics and the predicted values.
Furthermore, an optimum loss function should be built to offer the ideal model parameters based on the features of the forecasting error [33].

Appropriate forecasting models with suitable parameters help to provide reliable projections .

2.3.2.3 Configuration of Weather Forecasting models

Model configuration, such as the type of kernel function and the corresponding kernel parameters in kernel-based prediction models, the number of layers and hidden nodes, learning rate and optimizer choice in traditional neural networks and DNNs, have a significant influence on forecasting efficiency in practice.

A suitable model setup can direct the model's learning of optimal model parameters, resulting in improved forecasting performance [34].

Forecasting models are typically based on two primary stages: data pre-processing and parameters optimization.

The original data should first be pre-processed using techniques like feature scaling or signal processing, since gathered data might contain unidentified noise or outliers.

After that, significant features should be identified.

Then, forecasting models are developed and trained using the chosen features to discover the connection between the input characteristics and the desired outputs.

After that, the implemented models best parameters combination should be chosen based on their prediction efficiency of the available datasets.

Finally, using the trained models, weather predictions should be developed based on the best model configuration obtained.

DL has been widely used in the generation and application of alternative energies, particularly in the process of wind speed prediction, owing to its potent capabilities.

Despite the lack of understanding of atmospheric physics, artificial intelligence strategies such as deep learning and machine learning generate satisfactory wind speed forecasting results.

2.4 Renewable Energy

Energy crisis and climate changes had led to a remarkable increase in the production of green power supplies that initially comprises the solar and wind energy generation [35].

Various meteorological data sets can be extensively used at various stages of renewable energy projects. For instance, a precise site evaluation is required before constructing a solar or wind farm to guarantee the project's financial viability. Real-time observations and precise wind or solar radiation forecasts are needed during the operational phase of the established solar or wind energy site [36]. As a result, numerous diverse data sets are needed.

2.4.1 The Importance of Accurate Weather Prediction for Power Generation

The access to electricity is a critical social and economic indicator that has been steadily rising in recent years [37], but in revenge the global energy demand increased rapidly to consistently overtaken the available worldwide supply (Figure 2.5) [38].

The challenge that the authorities face now is the production of sufficient amount of energy, while reducing the carbon emission that significantly affects the planet being the primary driver of global climate changes.

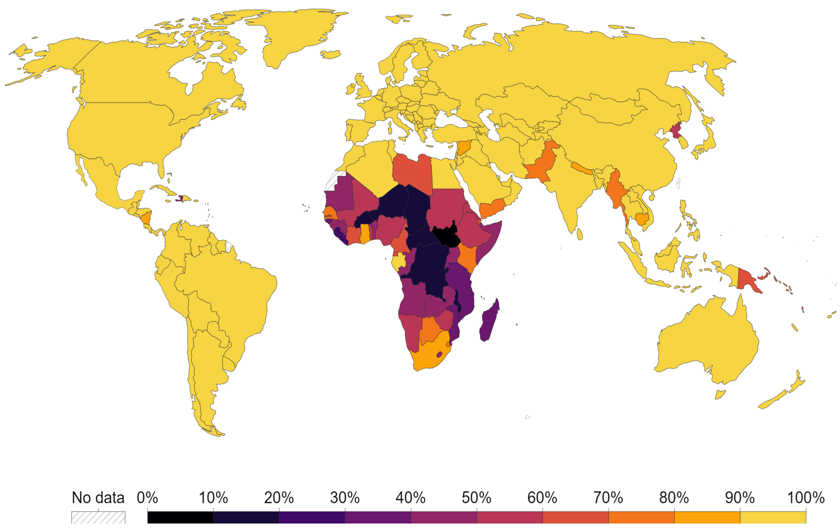


FIGURE 2.5: Electricity access.

Algeria's electricity consumption is rated as high compared to countries that have a similar development level [39].

Due to the remarkably expanding population, the country's total energy consumption grew with approximatively 5% per year over the past 10 years with a small perturbation between 2020 and 2021 (Figure 2.6).

Non conventional sources like wind, solar and geothermal energies have become the center of attraction and gradually transiting into the leading sources of power in the future.

Among the different renewable power generating systems, the wind power is considered as one of the largest shares of overall renewable growth in the past years.

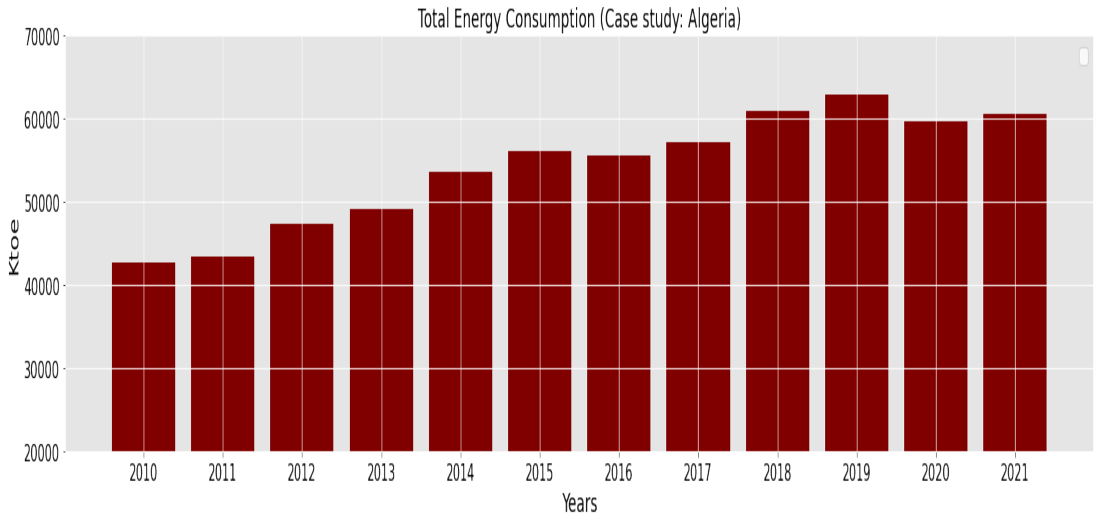


FIGURE 2.6: Total Energy Consumption (Case study: Algeria).

Algeria today is at a very critical stage in its energy transition strategy. The nation is well positioned and has both the potential and economic incentives to play a major role to maximize its natural resources, embrace clean power, and become a renewable energy leader .

In order to reach these expectations, the government has launched ambitious projects to enhance the capacity of the renewable energies manufacturing. From constructing wind power farms photovoltaic plants to hybrid power generation systems.

The country is also member of the German Desertec Industrial foundation, that promotes the acceleration of the energy transition by exploiting Sahara solar and wind power from the unproductive lands in order to supply 15 % of the electricity needs by 2050 .

Algeria is considered as a rich country that has promising wind energy potential of about 35 TWh/year. Nearly 50 % of the country consists of regions that dispose of significant wind speed levels.

The country's first and only wind farm is set at Kaberten zone in Adrar city with installed capacity of 10.2 MW funded from state-utlity Sonelgaz[40].

An enticing policy has been suggested by the Algerian government to promote investments in the wind power generation, but the hardest step remaining is to find the most suitable sites to reach the target of establishing 5(GW) of wind power by 2030 (Figure 2.7).

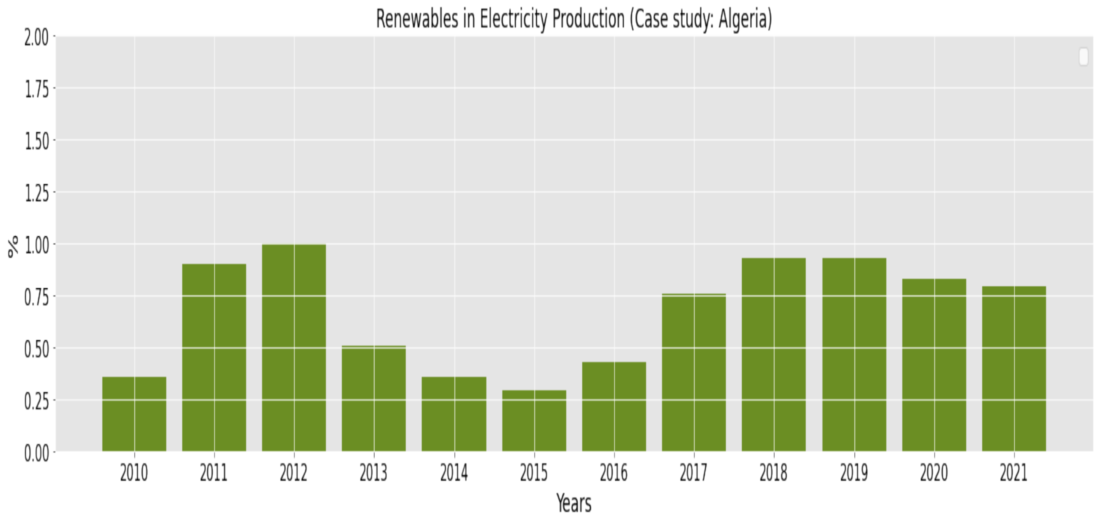


FIGURE 2.7: Renewables in Electricity Production (Case study: Algeria).

The amount of energy produced by wind turbines is mainly influenced by the wind speed. Stronger winds allow the generation of more power, because higher wind speed levels permit the acceleration of the turbine blades rotation that generates more mechanical power converted to more electrical energy .

2.4.1.1 Wind Energy Formula

There is cubic relationship between the wind power and the wind speed [34], determined by the following formula:

$$P = \frac{1}{2}\rho A V^3 \tag{2.11}$$

Where:
P = Wind power (W),
ρ = Density of dry air (kg/m^3),
A = swept area of blades given by :

$$A = \pi r^2. \tag{2.12}$$

Where:
r : radius of the blades (m).
V = velocity of the wind (m/s).

This relationship makes the wind speed such crucial key for wind energy production.
For this reason our efforts in this work are focused on forecasting the wind speed variable being the fundamental feature for the wind power generation.

2.4.1.2 The interest of Accurate Wind Assessments

The wind power generation is considered as an extremely sophisticated process that requires high preciseness and reliability of inputs and results.

Building a robust wind speed forecasting framework capable of handling the challenging fluctuated behaviour of the wind and its complex nature is highly recommended to reach the desired results.

Accurate wind speed forecasting improves the reliability of the wind energy generation task, and decreases the need of additional balancing and integrating energy reserves, thus, reducing the power generation production cost.

Below, a detailed table of some relevant works on wind power and wind speed prediction based on DL strategies.

A selection of relevant weather and wind forecasting frameworks that adopt a hybrid strategy in the forecasting process for renewable energy purpose is presented in Table 2.1.

Table 2.1: Summary Review of Hybrid Forecasting Models

Works	Time Scales	DL Models	Inputs	Outputs	Weights optimiz-technique	Data pre-processing	Single / Multiple-step
[41]	Medium-Range	LSTM-EFG	Wind power Wind speed	wind power	-CSO-Algorithm	EEMD technique	Single-step and Multiple-step-ahead prediction
[42]	Ultra-Short -Range	LSTM / two-layered GRU	Turbine state, rotation rate, generating capacity, sine of wind direction, cosine of wind direction, pitch angle, wind power.	Wind power	wind speed correction process	/	Single-step and Multiple-step-ahead prediction

[43]	Short-Range	1D-CNN + Bi-LSTM Vs FNN, 1D-CNN, LSTM, Bi-LSTM	2m temp, 2m relative hum, 10m wind speed	2m temp, 2m relative hum, 10m wind speed	/	-linear interpolation. -Data Reshaping -One-Hot Encoding -Max-min normalization	Single-step-ahead prediction
[44]	Short-Range	LSTM	Wind power	Wind power	/	-DWT -z-score normalization	Single-step and Multiple-step-ahead prediction
[45]	Medium-Range	DFF,DCN, RNN, Attention mechanism and LSTM	Wind power temperature	Wind power temperature	/	-Discrete wavelet -FFT -Min–max normalization	Multiple-step-ahead prediction
[46]	Short-Range	CNNGRU-SVR	wind speed	wind speed	Grid search algorithm	-SSA-	Signle step and Multiple-step-ahead prediction

[47]	Medium-Range	DBSCAN-SDAE-LSTM	wind speed, wind direction, pressure, Temperature, humidity	Wind speed	-ADAM-Algorithm	DBSCAN+ SDAE + Batch normalization	Single-step-ahead prediction
[48]	Short-Range	bidirectional GRU	wind speed wind direction	wind speed wind direction	-Adam-Algorithm	/	Single-step-ahead prediction
[49]	Short-Range	-LSTM-ELM-	wind speed	wind speed	/	VMD-SSA	Single-step and Multiple-step-ahead prediction

2.5 Conclusion

In this chapter, we have discussed the fundamentals behind the three main research axises.

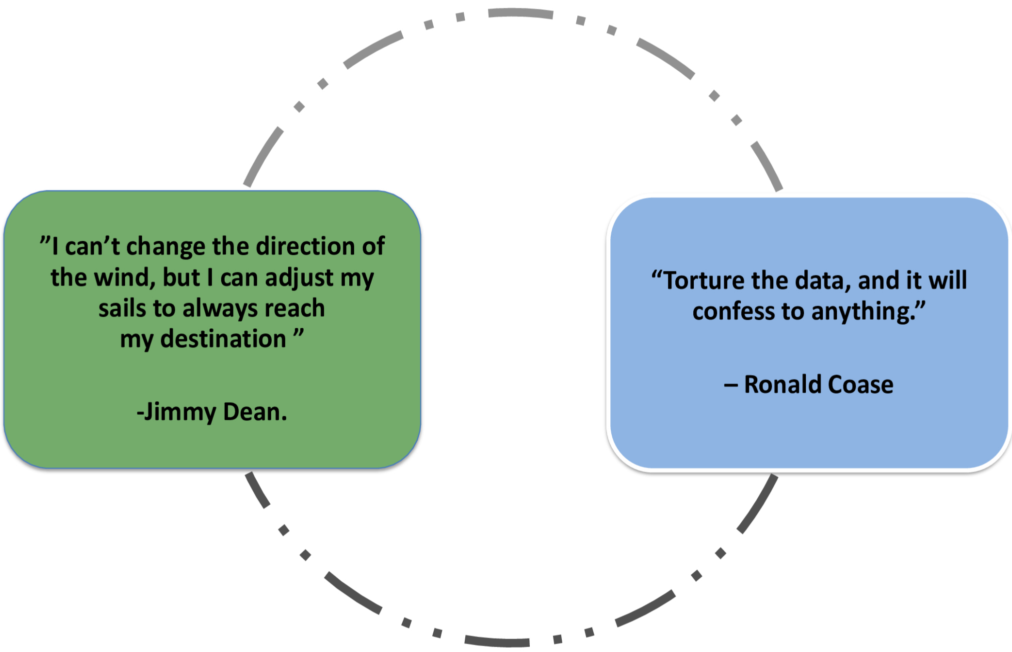
Starting by the theories behind the traditional weather forecasting strategies, with a brief description of some important weather factors that tend to change over time, generating time series of each parameter, where a forecasting model using this time series data should be developed.

The deep learning forecasting models are then explained being the best alternative for time series prediction, with the ability of automatically learning the temporal dependencies. A detailed explanation of the performance evaluation and model configuration is provided for the best model selection.

Finally a renewable energies section is dedicated, to better understand the process for an efficient renewable energies production (wind energy) based on weather forecasting parameters (wind speed) and the importance of accurate predictions for efficient electricity generation.

Chapter 3

Wind Speed Forecasting



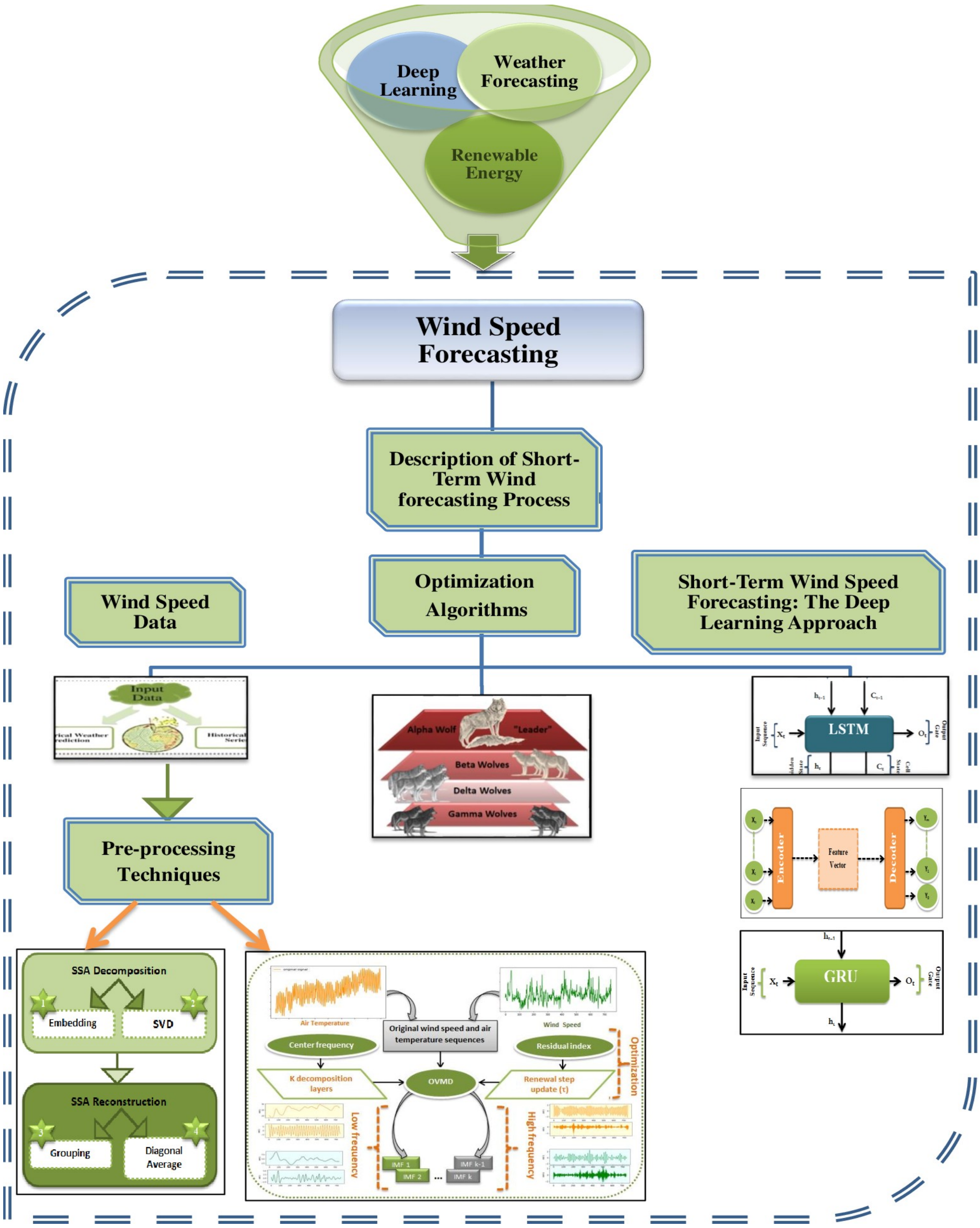


FIGURE 3.1: Chapter3 Map.

3.1 Introduction :

With 93.6 GW of fresh established capacity, 2021 was the second-best year on record for the worldwide wind industry, and it will continue to expand and contribute significantly in the global energy conversion toward a wealthy low-carbon economy.

Wind power not only offers clean, emission-free energy, local industry, and jobs opportunities, as well as attracting both domestic and foreign investment, but it also serves as an industrial lifeline for rural towns worldwide. New regions throughout the world are currently making significant leaps toward achieving the full potential of wind energy[50].

The increased usage of wind for renewable energy production, particularly in the power industry, raises the demand for system flexibility. Generation is no longer exclusively dependent on consumption, consumption also responds to changes in generation.

The capacity of energy that a wind turbine is capable of producing is highly dependent on the wind speed factor.

Today's wind speed databases are massive and collected from heterogeneous sources, that make them extremely vulnerable to noisy, missing, and inconsistent data. Low-quality data generates poor results and high-quality data produces more accurate predictions, so data preprocessing has become a critical step in the wind speed prediction process.

However, the complex and unpredictable nature of wind speed makes the forecasting process a challenging task to handle by numerical models, that is why the use of DL approaches is highly recommended to achieve accurate prediction results.

In this chapter, we will outline a selection of the most popular techniques and models for wind speed forecasting using data preprocessing methods, optimization algorithms, and deep learning models, with some relevant works on each one them.

3.2 Description of Wind forecasting Process

3.2.1 Wind Speed Forecasting over Time Horizons

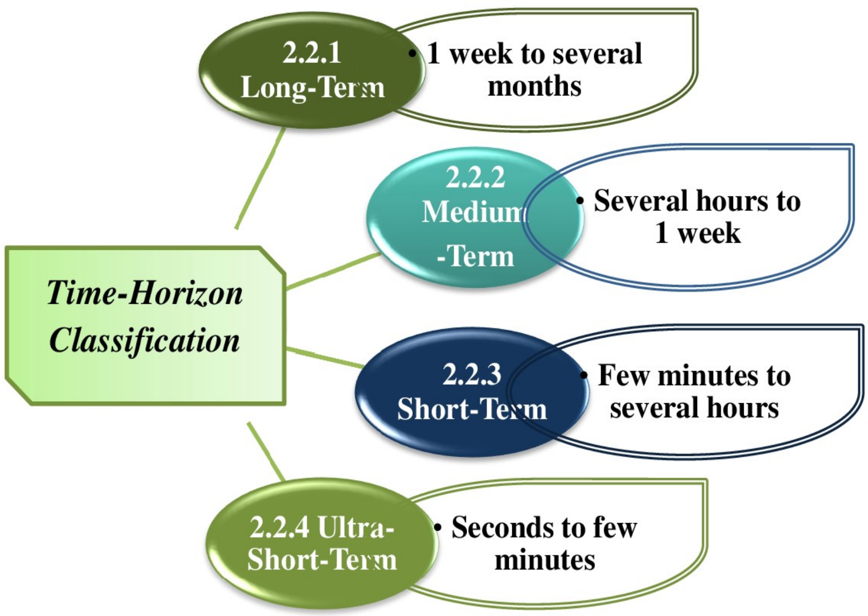


FIGURE 3.2: Time-Horizon Classification for Wind Speed Forecasting.

3.2.1.1 Long-range Wind Speed Forecasting

Long term wind speed forecasting is used to :

- Control long term operations of the wind power production systems.
- Maintenance scheduling of the wind farms.
- Managing operations cost

Wind Variability Source

- Climatic and seasonal changes.

3.2.1.2 Medium-range Wind Speed Forecasting

Medium-term wind forecasting helps in :

- The Estimation of the extra reserve requirements.
- Efficient market operation for wind energy trading.
- Generating online offline decisions.

Wind Variability Source

- Changes in weather patterns from the day and night cycle [51].

3.2.1.3 Short-range Wind Speed Forecasting

This type of wind forecasting is made for:

- Monitoring the load system balancing requirement.
- Contributing to the overall uncertainty of the power generation system.
- Reasonable load forecast decisions.

Wind Variability Source

- Thermal exchange between the atmosphere and ground [52].

- Short term forecasting is the origin of the physical principle of wind power generation.

3.2.1.4 Ultra-Short-range Wind Speed Forecasting

This type of wind forecasting is applied in:

- Turbine regulation and inspection.
- Real-time scheduling adjustment plans.

Wind Variability Source

- The flactuated nature of the wind speed [53].

3.2.2 Wind Speed Data :

Different wind speed information can be used in multiple stages of the wind power generation process.

For instance, accurate predictions and instantaneous data readings are highly recommended where the optimum dataset is mostly related to the application [54].

There are two sources for the wind speed input data as demonstrated in Figure 3.3 :

- Numerical Weather Prediction (NWP) data.
- Historical Time Series data.

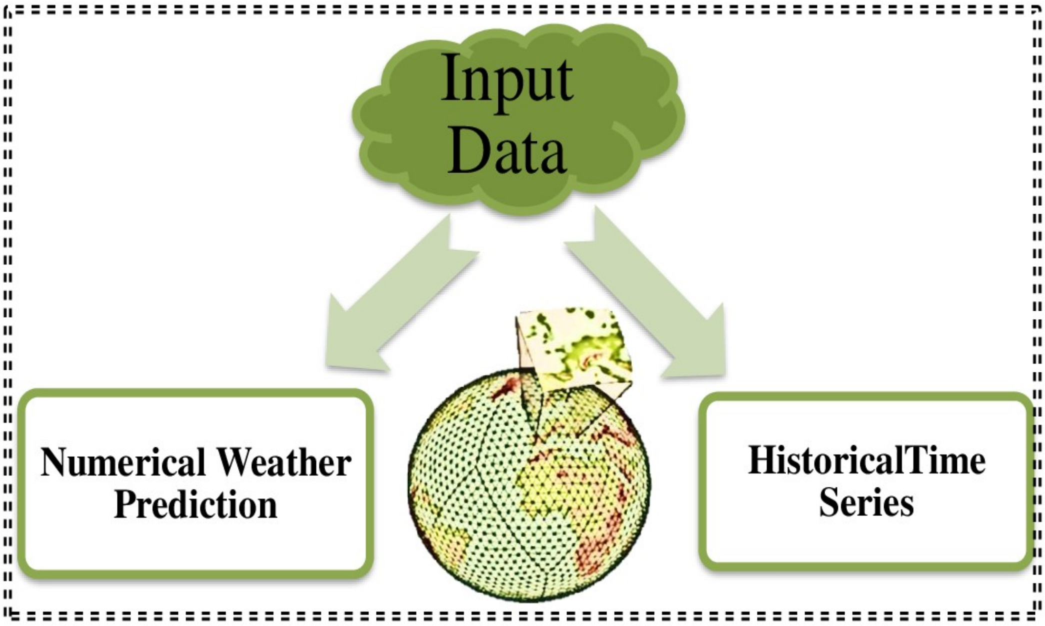


FIGURE 3.3: Wind Speed Input Data.

NWP data are generally used in Long-Term wind speed forecasting, while forecasting up to one day ahead can be performed using historical time series data alone.

Uncertainty of the measured wind speed data

Due to multiple metrics including hardware deficiency and reliability, additional meteorological factors, limited available data, and complexity of the wind nature, the collected data can be exposed to different uncertainties being : missing values, noise, redundant, and outliers.

This uncertainties may influence the forecasting process and prevent from training accurate forecasting models. Therefore, the pre-processing of the collected wind speed data is a key phase before conducting the wind speed forecasting process[55].

Features Extraction

Extracting meaningful patterns becomes increasingly important in wind speed prediction. It is mainly used to identify key features, summarize most of the information and effectively reduce the amount of data that must be processed, with the preservation of the original dataset characteristics[56].

It consists of transforming the wind speed time series into numerical features that can be processed later on, using different forecasting models in order to achieve better results.

Therefore, using accurate input features not only promotes efficient wind speed forecasting, but also helps the model to train more accurately and enhances the learning speed.

3.2.2.1 Wind Speed Data Pre-processing Techniques :

There are multiple data preprocessing techniques, each specialized in a particular data problem. These techniques are not mutually exclusive [57].

We can cite:

a. Data cleaning:

Used in noise filtering and also to correct the data incompatibility, as illustrated in Figure 3.4.

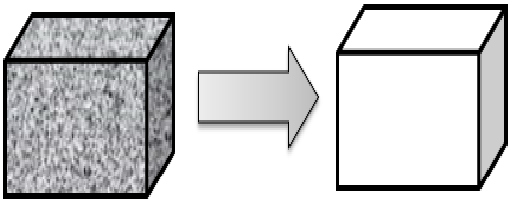


FIGURE 3.4: Data cleaning.

b. Data integration:

Combines information from separate origins into a coherent global dataset, as in Figure 3.5.

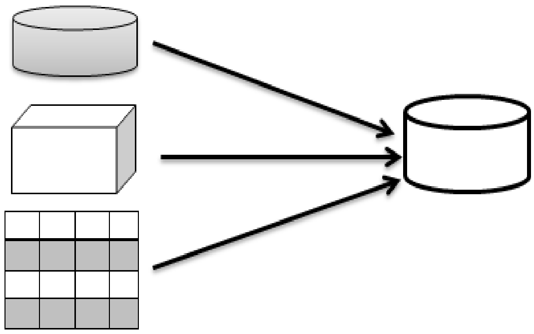


FIGURE 3.5: Data integration.

c. Data reduction:

The data size reduction can be made using aggregating, redundant features elimination, or clustering, as demonstrated in Figure 3.6.

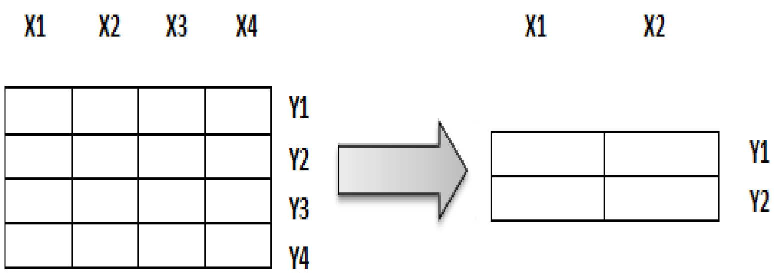


FIGURE 3.6: Data reduction.

d. Data transformation:

Such as normalization, used to scale the data in order to fall within a smaller range for example : from 0.0 to 1.0,[58]. As shown in Figure 3.7.

-5, 25, 18, -24, 100 → -0.05, 0.25, 0.18, -0.24, 1.00

FIGURE 3.7: Data transformation.

Several techniques have been proposed attempting to deal with the fluctuated and complex nature of wind speed data. They can be organized into many classes, which includes the Feature scaling and signal processing class[59].

3.2.2.1.1 Feature Scaling :

- Rescaling (min-max normalization) :

A commonly used normalization method that converts the minimum value of each feature to 0 and the maximum value to 1. The rest values get transformed into decimals between 0 and 1[60].

- Standardization (Z-score/standard score Normalization):

A normalization technique to standardize scores on an identical scale, dividing each score deviation by the standard deviation in a dataset.

The outcome is a standard score, where values less than the mean presents negative z-score, and positive scores designates the values greater than the mean. The average of every z-score for a dataset is zero [61].

- The Batch normalization :

Standardization technique of the inputs of a layer which allows the acceleration of the training process in deep learning networks, and enhances the performance of the model [62].

3.2.2.1.2 Signal Preprocessing :

The collected wind speed data is usually treated as a signal.

Meteorological researchers have employed signal processing techniques to decompose wind data to multiple subseries.

Then, the resulted subseries are forecasted separately and combined to generate the final wind speed forecast, based on the original data.

Signal processing techniques can also be used for smoothing and removing noise from the collected wind speed information.

Smoothing wind speed datasets is the removal of random changes that appear as coarseness in a raw time series data visualization. It eliminates noise to highlight the signal, that may contains patterns and cycles.

The resulted denoised data is then maintained for training the dedicated prediction strategies[63].

Five types of wind speed signal pre-processing techniques are commonly used for smoothing and decomposing the wind speed original data, that we investigated in detail in this chapter, being: Moving Averages, Fourier Transforms, Wavelet Transforms-based techniques, Mode Decomposition-based and Singular Spectrum Analysis (SSA)-based methods.

These techniques are shown in Figure 3.8, and Table 3.1 [52].

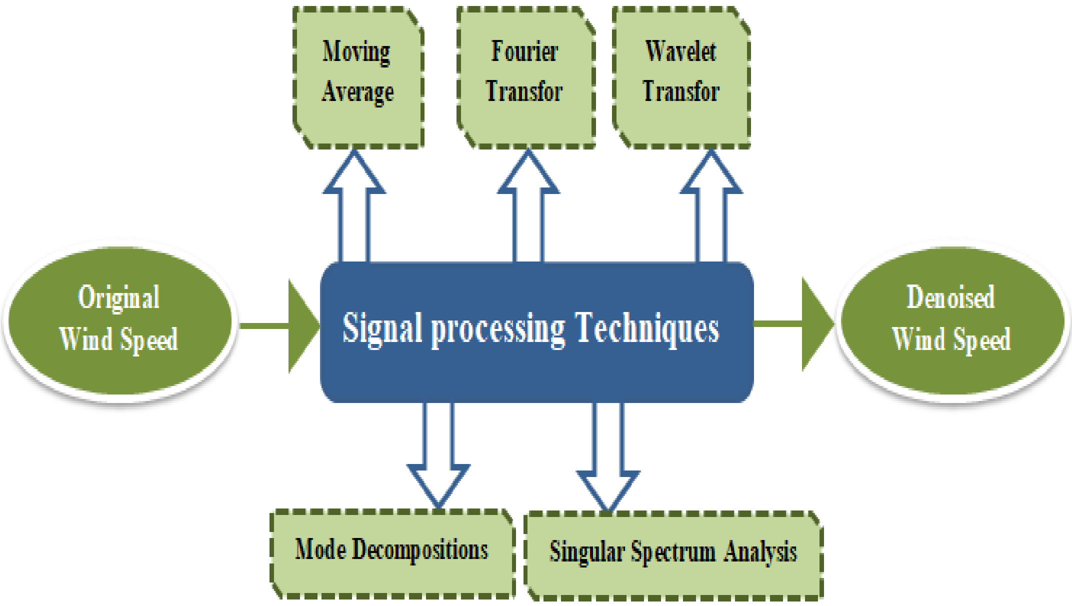


FIGURE 3.8: Signal Preprocessing Process.

Table 3.1: Related Wind Speed Signal Pre-processing Techniques.

Type	Authors	Advantages	Inconvenient
Moving Average	**Ellis et al** 2005[64] **Cadenas et al** 2010 [65] **Cambron et al** 2016[66]	-TMA: Noise reduction and hidden patterns reveal. - EMA : Timely and alleviate the negative impact of lags. WMA: Underperformances and small shifts detection.	- Generally slower to respond to changes in the dataset. - The entire data history must be carried along with the computation. - Hidding the actual distribution if the data lake normal distribution.
Fourier Transforms	**Oh et al** 2018[67] **Jaseena et al** 2021[68]	- DFT : Completely discrete both in frequency and time - FFT: Determine the frequencies and the amplitude at each f.	- Aliasing effects, amplitude modulation spectrum -Restricted range of waveform data.
Wavelet Transforms	**Debnath et al** 2017[69] **Brusco et al** 2022[70] **Jnr et al** 2022[71]	-DWT: Solving the time lag problem. -DWT: Extract local spectral and temporal information -CWT: Good to find damping ratio of oscillating signals	-Shift sensitivity and poor directionality -Depends on the choice of the wavelets - Adds excess redundancy - Computationally intensive
Mode Decompn	**Huang et al** 2019[72] **Hu et al** 2021[73]	- OVMD: Reduction of the wind patterns dimension -VMD: Determination of relevant bands adaptively	-Lacks a comprehensively verification -Suffers from the limitation of the Fourier spectrum
SSA	**Liu et al** 2018[74] **Mi et al** 2020[75]	-SSA: can be generalized to many time series. -Identify similar spectrals in multiple time series	- Imitated when it processes non-stationary multi-components signals

a. Moving Averages :

Moving average (MA) is a traditional approach for decomposing data. The theory behind moving average is that consecutive observations usually have similar values. For this, using the averages of nearby observations allows to reach a coherent estimation of the trend at the required state.

Using the observation's averages grants the elimination of most of the randomness in the data, leaving a smooth trend results.

This method forms the basis of most time series decomposition approaches and it is divided to three types: Simple, Exponential and Weighted Moving Averages[76].

a.1. Simple Moving Average :

The simple moving average (SMA) calculates the mean of the last n observations, where Ox is the number of observations in an interval of time, and n is the number of observations (Eq.3.1){[64].

$$SMA = \frac{O_1 + O_2 + O_3 + ... + O_x}{n} \tag{3.1}$$

There are two types of SMA :

a.1.1. Trailing Moving Average(One-sided moving average):

For each average, Trailing Moving Averages (TMA) makes use of the current and prior observations.

For instance, the maths behind TMA of O at time t with a length of n is:

$$TMA_n = \frac{O_{t-(n-1)} + O_{t-(n-2)} + O_{t-(n-3)} + ... + O_t}{n} \tag{3.2}$$

Trailing moving average is used for time series forecasting and mainly uses past observations.

a.1.2. Centered moving averages (two-sided moving average):

To determine the average at a specific moment in time, centered moving average (CMA) uses both past and future data that surround it.

The formula of CMA of O at time t with n length is:

$$CMA_n = \frac{O_{t-(n \div 2)} + O_{t-(n \div 2-1)} + ... + O_t + O_{t+1} + ... + O_{t+(n \div 2)}}{n} \tag{3.3}$$

Since centered intervals permit an equal number of observations before and after the moving average, they work out equally for an odd number of observations. When the length is even, the computations must be adapted by using a weighted moving average (Eq. 3.4).

$$CMA_n = \frac{(0.5 * O_{t-(n \div 2)}) + O_{t-(n \div 2-1)} + ... + O_t + O_{t+1} + ... + (0.5 * O_{t+(n \div 2)})}{n} \tag{3.4}$$

The simplest form of moving averages is the simple SMA.

a.2. Exponential Moving Average (EMA):

In order to make the calculation more sensitive to recent observation, the EMA accords more weight to recent observations. The EMA is a very popular scheme that generates a smoothed Time Series [65].

To calculate the EMA:
1- First, we need to compute the (SMA) for the given period.
2- Then, compute the ”smoothing factor”, which is the multiplier for weighting the EMA being:

$$Smoothingfactor = \frac{2}{selectedtimeperiod + 1} \tag{3.5}$$

3- The formula for EMA is given as follows:

$$EMA = O_t \times Smoothingfactor + SMA_y \times (1 - Smoothingfactor) \tag{3.6}$$

where:
O = Observation
t = Today.
y = Yesterday.

There are two extensions of the EMA being: Double Exponential Smoothing which is more effective in dealing with trends, and Triple Exponential Smoothing that performs better when dealing with parabolic trends.

c. Weighted Moving Average (WMA):

WMA gives more weight to the more recent observation points since they are more relevant than data items from the distant past.

The total of the weightings should equal 1 (or 100%) [66].

The formula of the WMA is giving as follows:

$$WMA = \frac{O_1 \times n + O_2 \times (n-1) + ...O_n}{\frac{n \times (n+1)}{2}} \tag{3.7}$$

Where:
n = Time period.

The WMA is more customized than the SMA and EMA. Weights applied to WMAs can differ based on the number of periods used in the computation.

b. Fourier transforms (FT) :

Fourier analysis transforms the time series from its original domain to a frequency component.

Instead of a function of time, the FT extracts frequency information embedded in time series to get smoother and easier data to be processed by the forecasting model, while reducing the number of the calculations required to the final forecast[77].

The FT transforms the signal into sines and cosines, that can be classified into different types including:

b.1. Fourier Series :

If the function f(x) is periodic, the expression of f(x) as a sequence of frequency terms with variable terms can be accomplished with discrete frequencies, with an endless number of terms (Eq.3.8).

$$f(x) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} a_n cos\ nx + \sum_{n=1}^{\infty} b_n sin\ nx \tag{3.8}$$

where,

$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)dx$
$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)cos\ nx\ dx$
$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)sin\ nx\ dx$ n $= 1, 2, 3 \ldots.$

- For an even function, only cosine terms exist.

- For an odd function, only sine terms exist.

b.2. Fourier Integral :

The Fourier Integral of f(x) on the interval $(-\infty, \infty)$ is given by:

$$f(x) = \frac{1}{\pi}\int_0^{\infty} A(\lambda)cos(\lambda x)d\lambda + \frac{1}{\pi}\int_0^{\infty} B(\lambda)sin(\lambda x)d\lambda \tag{3.9}$$

where,

$$A(\lambda) = \int_{-\infty}^{\infty} f(t)cos(\lambda t)dt \tag{3.10}$$

and

$$B(\lambda) = \int_{-\infty}^{\infty} f(t)sin(\lambda t)dt \tag{3.11}$$

For the non periodic functions, integrals with limits of infinity are applied[78].

b.3. Discrete Fourier Transform :

The Discrete Fourier Transform (DFT) is the Fourier Transform required for time series data when the input data is recorded at discrete intervals (Eq.3.12).

$$X_k = \sum_{n=0}^{N-1} x_n \cdot e^{-i2\pi kn/N} = \sum_{n=0}^{N-1} x_n[cos(2\pi kn/N) - i \cdot sin(2\pi kn/N)] \tag{3.12}$$

where:

X_k = The DFT which includes information of both amplitude and phase.

N = number of samples

n = current sample

k = current frequency, where k[0,N1]

x_n = the sine value at sample n

These equations are impracticable since the number of calculations is proportional to N^2, it is impossible to deliver results in a reasonable period even using a powerful computer.

The DFT is made efficient for computing via the Fast-Fourier Transform (FFT) technique [67].

d.4. Fast Fourier Transform :

The Fast-Fourier Transform (FFT), often recognized as one of the most important algorithms of the twentieth century, is exactly what puts the Fourier Transform concept into practice .

The FFT is a fast algorithm for calculating the DFT.

The FFT successfully decreases the DFT's complexity from O(n^2) to O(nlogn). with n as the data size.

This decrease in calculation time is considerable, particularly for data with important N.

The process for FFT smoothing algorithm can be described as follows:

1. The calculation of the FFT from the input signal of wind speed.

2. The frequency domain progression of the modified data.

3. The removal of high frequency patterns in order to generate a smoothing effect.

4. Inverse FFT for the reconstruction of the original signal [79].

The FT has a significant drawback because of the global frequency data collection, being the frequencies that persist over the whole signal.

This category of signal decomposition may not be suitable for domains where signals include short intervals of distinctive oscillation.

The Wavelet Transform is a more advantageous alternative method that decomposes the function into a combination of wavelets localized in both frequency and time.

c. Wavelet-Based Approaches :

The Wavelet Transform (WT) is an improved method for evaluating signals that have a dynamical frequency spectrum in both the frequency and temporal domains, while offering a high resolution.

Working with different scales the WT shows which frequencies are present in a transmission, and when these frequencies occur [69].

The WT is a mathematical approach for decomposing a signal into numerous lower resolution levels by manipulating the scaling and shifting parameters of a single wavelet function.

The WT is an endless collection of different transforms, depending on the merit function used to compute it. There are several methods for sorting the wavelet transform types.

We can create discrete wavelet transforms using orthogonal wavelets and continuous wavelet transforms using non-orthogonal wavelets.

c.1. Discrete Wavelet Transform :

The Discrete Wavelet Transform (DWT) is a wavelet transform implementation that employs a discrete combination of wavelet scales and translations.

This transform decomposes the signal to reach a mutually orthogonal combination of wavelets. It proved its efficacy and robustness in removing high frequency noise.

DWT can be performed easily and effectively to denoise a flactuated signal, and getting more or less denoised signal using only a restricted set of highest coefficients from the DWT spectrum, and applying an inverse transform (with the same wavelet basis)[70].

DWT is usually used as a filter-bank. This indicates that it is constructed as a series of high-pass and low-pass filters.

The Gwyddion, adaptive thresholding, scale adaptive thresholding, and scale-space universal thresholding are usually the approaches used for choosing the coefficients that will be maintained.

To find the threshold within these approaches, we first :

1. Compute the noise variance using the finest wavelet coefficients given by :

$$\sigma = med(|Y_i j|)/0.6745. \tag{3.13}$$

Where:

$Y_i j$ refers to coefficients of the decomposition's high scale subband (where the majority of the noise is expected to exist).

2. Calculate the limit of the noise threshold (N : the signal length) :

$$\lambda = \sigma * sqrt(2 * ln(N)). \tag{3.14}$$

3. When a particular scale's threshold is identified, we may either eliminate all coefficients with absolute values less than the threshold value (hard threshold), or decrease the absolute value of these coefficients by the threshold value (soft threshold).

There are several discrete wavelet families. Each one chooses a distinct trade-off in how compact and smooth the wavelet looks.

The selection of the wavelet family depends on the characteristics of the signal that best matches.

Below in the table we define some DWT families with their properties.

TABLE 3.2: Some Discrete Wavelet transform Families

DWT Family	Properties
Daubechies (db)	Asymmetric, orthogonal, biorthogonal, Nonlinear phase, energy concentrated near the start of their support, highest number of vanishing moments N for a given support width.
Haar (haar)	Asymmetric, orthogonal, biorthogonal, special case of Daubechies , useful for edge detection.
Coiflets (”coif”)	Near symmetric, orthogonal, biorthogonal, Scaling function and wavelets have same number of vanishing moments N.
Symlets (”sym”)	Least asymmetric, orthogonal, biorthogonal, nearly linear phase; N vanishing moments .
Biorthogonal (”bior”)	Symmetric, not orthogonal, biorthogona.

Choosing the wavelet is a critical phase for smoothing data. Orthogonal wavelet based denoising is known as the most effective method to enhance the quality of a signal, in particular Daubechies wavelet family, which is considered as the best choice for denoising time series signals such as wind speed data.

c.2. Continuous Wavelet Transforms :

Continuous wavelet transform (CWT) is a wavelet transform implementation which employs variable scales and almost arbitrary wavelets (Eq.3.15).

The wavelets applied are not orthogonal, and the resulting data is strongly correlated.

This transform can not be used for discrete datasets, unless the finest wavelet translations are equivalent to the data sampling.

This is also called the Discrete Time Continuous Wavelet Transform (DT-CWT), and it is the usually employed technique of computing CWT in real-world domains[71].

$$X_w(a,b) = \frac{1}{|a|^{1/2}} \int_{-\infty}^{\infty} x(t)\psi^*(\frac{t-b}{a})dt \tag{3.15}$$

where, ψ is the continuous mother wavelet translated by a factor of b and scaled by an amount of a.

There can be an endless number of wavelets since the scaling and translating variables values are continuous.

In general, the continuous wavelet transform functions by simply applying the wavelet transform theory, i.e. performing a sort of convolution to the signal using the scaled wavelet.

Generating an array of the same length N as the signal for each scale. Using M randomly defined scales, a field $N * M$ is created that directly describes the time-frequency plane.

This calculation algorithm generally lays on direct convolution or a convolution through Fourier multiplication in Fourier space so called Fast Wavelet Transform (FWT)[69].

In the table bellow a detailed comparison between the DWT and CWT is presented.

Table 3.3: Comparison between Continuous and Discrete Wavelet Transforms

Properties of DWT	**Properties of CWT**
DWT employs exponential scales with base equivalent to 2 .	CWT employs exponential scales with bases less than 2.
The generation of fewer coefficients makes the DWT more computationally efficient.	The important number of coefficients, occupy considerably the RAM, and slow down the execution speed.
DWT can be implemented using filter banks	CWT can not be performed using filter banks.
Signal Denoising,feature extraction and compression process.	Time-frequency analysis, and signal transient location.

d. Mode Decomposition-Based Methods

The purpose of using signal decomposition methods is to separate and extract signal components from complex signals. It is also a useful way for identifying modal information in time-domain signals

There are different types of mode decomposition techniques, including :

The Empirical Mode Decomposition (EMD) versions : Ensemble EMD (EEMD), Fast EEMD (FEEMD), and Complete EEMD (CEEMD).

In this section we will focus on a similar recent alternative being the Variational Mode Decomposition (VMD) with it's two variants , the Adaptive VMD (AVMD) and the Optimal VMD (OVMD)[34].

d.1. Variational Mode Decomposition :

Variational Mode Decomposition (VMD) is considered as one the most recent signal decomposition tools. The VMD was implemented by Dragomiretskiy et al[80].

This technique efficiently overcomes the aliasing modal challenge of EMD versions with improving the model noise tolerance, reliabilty, and computing economy. It has been broadly used for wind speed forecasting that may nonrecursively decompose the wind signal $f(t)$ into k distinct band-limited IMFs u_k (k=1,2,..,K).

There is a center frequency ω_k for each mode u_k. The objective of the decomposition is to reduce the total frequency bandwidth of each mode, requiring the mode's aggregation to be equal to the specified series $f(t)$.

This approach lays on the Hilbert transform, the traditional Wiener filtering, and the mixed frequency variational issue. It consists of two parts: building variational and solving variational problems, detailed below :

(1) Construction:

$$\begin{gathered} \overset{min}{U_k, \omega_k} \quad \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * U_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ s.t. \qquad \sum_k U_k = f \end{gathered} \tag{3.16}$$

where,
$\delta(t)$ is the Dirac distribution and $*$ denotes convolution.

(2) Solution:

Enhanced Lagrange function is generated by inserting the Lagrange multiplier λ(t) and the quadratic penalty factor α as follows:

$$\begin{aligned} L(\{u_k\}, \{\omega_k\}, \lambda) = \alpha \sum_k \left\| \partial_t \{ [\delta(t) + \frac{j}{\pi t}] * u_k \} e^{-j\omega_k t} \right\|_2^2 \\ + \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \langle \lambda(t), f(t) - \sum_k u_k(t) \rangle \end{aligned} \tag{3.17}$$

The Alternate Direction Method of Multipliers (ADMM), based on dual decomposition and the Lagrange method, is used to solve the non-constrained variational problem by iteratively updating u_k^{n+1},w_k^{n+1} and λ^{n+1}.

The FT of $u_k^{n+1}(t)$, $u_i(t)$, $f(t)$ and $\lambda(t)$ are $\hat{u}_k^{n+1}(\omega)$, $\hat{u}_i(\omega)$, $\hat{f}(\omega)$ and $\hat{\lambda}(\omega)$, respectively.

We can get the $\hat{u}_k^{n+1}$ using the inverse FT of the real part of $\hat{u}_k^{n+1}(\omega)$, while n is the iteration number[72].

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i<k} \hat{u}_k^{n+1}(\omega) - \sum_{i>k} \hat{u}_k^{n+1}(\omega)}{1 + 2\alpha(\omega - \omega_k^n)^2} + \frac{\frac{\hat{\lambda}^n(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k^n)^2} \tag{3.18}$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k^{n+1}(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k^{n+1}(\omega)|^2 d\omega} \tag{3.19}$$

The $\hat{\lambda}^{n+1}(\omega)$ is updated based on the ADMM using Eq.(3.20), t : the update parameter.

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau(\hat{f}(\omega) - \sum_k |\hat{u}_k^{n+1}(\omega)|) \tag{3.20}$$

The VMD iterative process is illustrated as follows:

Step 1: Given :

- Wind speed data $f(t)$.

- Modes initialization u_k^1.

- Center frequencies ω_k^1

- Lagrangian multipliers λ^1.

Step 2:

- The update of $\hat{u}_k(\omega)$, ω_k and $\hat{\lambda}$ for each mode u_k.

Step 3:

- The iteration is stopped, when the condition is satisfied.
-$u_k^{n+1}(t)$ can be obtained by inverse FT of $\hat{u}_k^{n+1}(\omega)$.

VMD decomposes the Fourier spectrum into segments to separate distinct components of the signal. This approach suffers from the Fourier spectrum limitation [81], when different components cannot be distinguished with the Fourier spectrum, they cannot be decomposed by VMD.

To address the drawbacks of the VMD approach, two novel decomposition extensions are introduced, called Adaptive Variational Mode Decomposition (AVMD) and Optimal Variational Mode Decomposition (OVMD).

d.1.1. Adaptive variational mode decomposition

The presented AVMD is known for its capacity of achieving an ideal parameters combination $[K, \alpha]$. The AVMD approach is very adaptable and noise-resistant.[82].

The AVMD was introduced to calculate the number of modes in an automatic way using the intrinsic mode function features. This approach evaluates the VMD decomposition findings using a set of indicators such as extreme value in the frequency level, permutation entropy, kurtosis criteria, and the loss in energy coefficient, among others.

Depending on the results, AVMD alter the mode number K and re-analyze the signal until the correct K value is reached. AVMD processes a variety of data, including noise-free signals and noisy signals..

d.1.2. Optimal Variational Mode Decomposition

As shown by the equations above in the VMD section, the number of decomposed modal functions K and the renewal updating step (τ) have a significant impact on the VMD's performance. An Optimal-VMD is then suggested by Huang [72], in order to optimize layer and update renewal step.

The central frequency technique (CF) and the residual index methodology are the basis of this novel extension.

The whole process of the OVMD is shown in Fig. 3.9.

- *Central Frequency method (CF) :*

The CF method is employed to compute the number of VMD decomposition layers K. It requires finding the mean of decomposition modes for various K values by expanding the input of the K layers beginning with K = 1. Every single time step, a K value is inserted and the CF method is computed. The resulting value is determined as the optimal K value after the frequency stabilization.

- *Residual Index method (REI) :*

The REI technique is introduced to calculate the value of (τ) (Eq.3.21). This approach depends on selecting the sequence with the lowest denoised data signal close to the original one.

$$REI = min\frac{1}{N}\sum_{i=1}^{N}\left[\sum_{K=1}^{K}U_k - f\right] \tag{3.21}$$

Where,

U : the decomposition modes number.

f : the original signal and N: the number of signals.

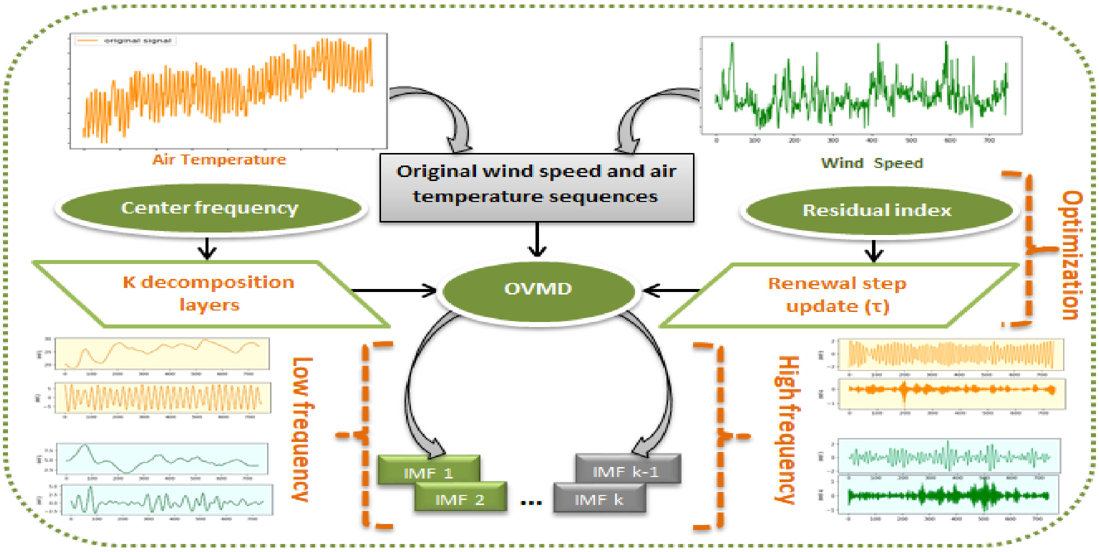


FIGURE 3.9: The Optimized-VMD flow-chart.

e. Singular Spectrum Analysis (SSA):

In recent years a powerful time series analysis approach called Singular Spectrum Analysis (SSA) has been introduced.

This technique provides a competitive smoothing effect and a remarkable enhancement in the prediction performance[53]. The SSA is a reliable time series preprocessing technique which is divided into two complimentary phases: decomposition and reconstruction (Fig.3.10).

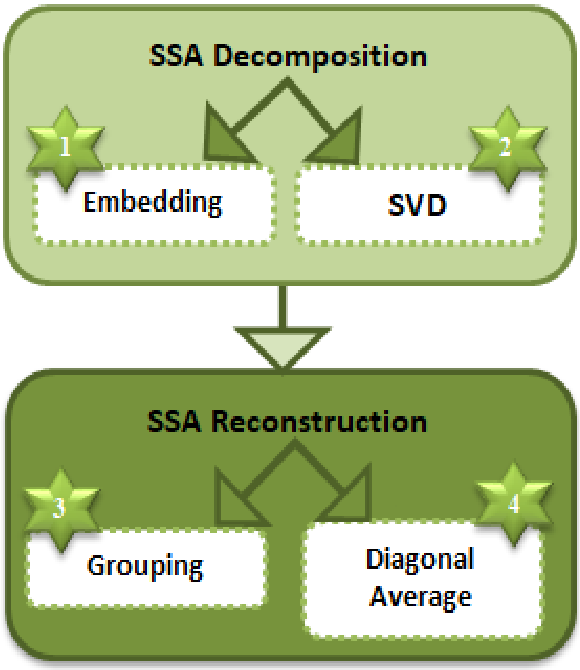


FIGURE 3.10: The SSA denoising architecture.

Where each has two phases explained bellow :

e.1. Decomposition :

- *Embedding :*

At this phase, the wind speed data, X = (X1.....Xn) is defined as the L-dimensional series and unfolded into a $Z = (Z_1.....Z_k)$, where $Z_i = (X_i, ...X_{i+L-1})$ a sequence of vectors in Eq.3.23.

Forming $K = N - L + 1$ lagged vectors.

L is the window length, included in $2 \leq L \leq N$.

$$Z = \begin{pmatrix} X_1 & X_2 & ... & X_k \\ X_2 & X_3 & ... & X_{k+1} \\ ... & ... & ... & ... \\ X_L & X_{L+1} & ... & X_N \end{pmatrix} \tag{3.22}$$

Where all elements of the trajectory matrix are equal in the diagonal.

- *Singular value decomposition (SVD) :*

After the embedding phase, we proceed the trajectory matrix employing the SVD [83], using the following formula :

$$Z = \sum_{i=1}^{d} Yi \tag{3.23}$$

Where,

d : the number of positive eigenvalues from the Eq.3.23 matrix.

$$Zi = Ui\sqrt{ZZ^T}Vi \tag{3.24}$$

Where,

V , U : the left and right eigenvector of the matrix ZZ^T, respectively

e.2 Reconstruction :

- *Grouping :*

The important components of the SVD are used to reconstruct the wind speed sub-layers, the least important components are considered as noise then removed.

The matrix Z is rewritten as follows:

$$Z = \sum_{i=1}^{m} Zi \tag{3.25}$$

Where;

$m \in [1, d]$, i varies from 1 to m.

$\{ Z_1, Z_2, ...Zm\}$ are the important components.

- *Diagonal averaging :*

In this final phase, each matrix Zi_j of the grouped decomposition is converted into wind speed sub-layers of length m. Using the diagonal averaging procedure, each matrix Zi of the $\{ Z_1, Z_2, ...Zm\}$ is transformed into new wind speed data series that are then reconstructed, as follows:

$$X = X_{I_1} + X_{I_2} + X_{I_3} + ... + X_{I_m} \tag{3.26}$$

Where $I_1, I_2, ...I_m = 1, 2, ...m$, and $\{ X_{I_1} + X_{I_2} + X_{I_3} +... + X_{I_m} \}$ are the reconstructed series of $\{Z_1, Z_2, ...Z_m\}$

The main key of the SSA process is the window length value , being the responsible of defining the embedding step.

Therefore, the selection of the appropriate window length is crucial to reach efficient denoising results. For this, the use of powerful optimizers is highly recommended to achieve the optimal parameters selection, and to obtain the desirable wind speed data smoothing effect [84].

3.2.3 Optimization algorithms :

The procedure of selecting a combination of optimal parameters values for a signal processing method or the best hyperparameters combination for a DL model is called, hyperparameter tuning.

A model parameter is known as a hyperparameter. Its value is defined before the start of the learning process. Hyperparameter tuning is considered as the cornerstone for deep learning algorithms.

The Optimization algorithms are the ones responsible of choosing the optimal parameters combination for the pre-processing techniques or the hyperparameter tuning of the forecasting models[85].

3.2.3.1 Grid Search Optimization Algorithm :

The Grid Search is an extremely effective optimization technique. It is most typically employed in deep learning models for hyperparameter optimization. It allows to find the optimum parameters for the most optimization issues from a list of parameters provided at the start, therefore automating the 'trial-and-error' process.

Grid search is known as the fundamental hyperparameter selection technique. Essentially, it partitions the hyperparameter into a discrete grid. Then, attempts all conceivable combinations of values from this grid, measuring various accuracy metrics, and selecting the architecture which produces the best results [86].

Grid search is an exhaustive method that searches all possible combinations to locate the best point in the domain. The main disadvantage is that it is time consuming. It takes a long time to check every combination, so it is mainly used when looking for the best combination of values of the hyperparameters.

3.2.3.2 The Adam Optimization Algorithm :

Adam algorithm was first introduced by by Kingma and Ba in 2015. It is generally applied in the optimization of time series problems .

This optimizer is an adaptive learning rate optimizer, that ussually replace the classical stochastic gradient descent procedure for training DL models, and handle sparse gradients on noisy problems. It uses adaptive learning rate approaches to determine individual learning rates for each parameter.

The Adam optimization algorithm is essentially a combination of momentum and RMSprop. It adjusts the weights of the network iteratively laying on training data and employing the squared gradients in order to scale the learning rate same as the RMSprop, taking advantage of momentum by employing moving average of the gradient in the place of gradient itself [87].

The Adam optimizer generates better results than the rest of optimization methods. reducing the calculation time, and depends on less tuning parameters.

As a result , Adam is defined as the most appropriate optimizer for the majority of DL applications.

3.2.3.3 Meta-heuristic optimization algorithms (Grey Wolf Optimizer):

The following section studies the Grey Wolf Optimizer (GWO) in depth.

The GWO is considered the most popular swarm intelligence optimizer. Because of its superiority over the commonly used swarm intelligence algorithms, it has been customized and adapted for multiple optimization applications.

This optimizer uses comparatively few parameters, without requiring derivation information in the first search. It is also not complicated, easy to employ, scalable, and adaptable. It is also known by the distinctive ability to achieve the proper equilibrium between exploitation and exploration during the investigation, this results in favorable convergence.

Figure 3.11 illustrates the GWO algorithm packs.

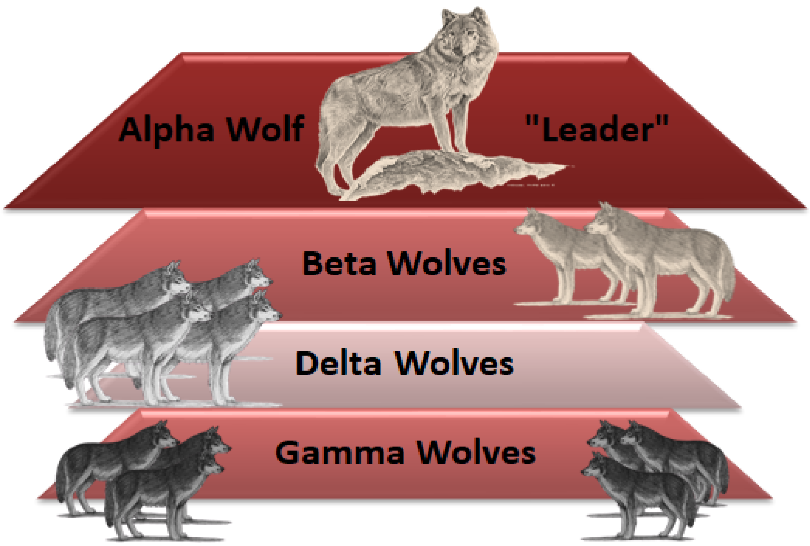


FIGURE 3.11: Grey Wolf Optimizer

The GWO is a meta-heuristic optimizer proposed by [88] , This algorithm simulates the intelligent social behaviors of the wolf packs. Where the fitness function is responsible of setting the amount of wolves. Based on the fitness optimum, the alpha wolf is the optimal fitness solution, who leads the pack in feeding, hunting and migration, reinforced by subordinate wolves known as the β and δ wolves, these resulted solutions are designated as fundamental-group. The remaining wolves are omegas.
The search for optimal solution process is guided by α, β, and δ wolves, followed by ω ones. The Grey wolves carry out a series of dynamic processes when hunting prey: pursuing, surrounding, harassing, and attacking. This enable the wolves to pursue a considerable number of preys.

To mathematically model the GWO behavior the upcoming equations for each phase are introduced:

a. Surrounding the prey :

The suggested encircling method defines a neighborhood in the shape of a circle centered on the solutions, that may be enlarged to greater dimensions as a hyper-sphere. At this stage, GWO algorithm examines a pair of points in n-dimensional space, then updates one of the points depending on the position of the second.

To model this, the following Eq. 3.27 and 3.28 have been suggested.

$$\vec{D} = |\vec{C} \cdot \vec{X}_p(t) - \vec{X}(t)| \tag{3.27}$$

$$\vec{X}(t+1) - \vec{X}_p(t) - \vec{A} \cdot \vec{D} \tag{3.28}$$

where,

X: GW position vector.
X_p: Prey position vector .
t: Current iteration.
A , C : Coefficient vectors.

A and C vectors are calculated as follows:

$$\overrightarrow{A} = 2\overrightarrow{a} \cdot \overrightarrow{r}_1 - \overrightarrow{a} \tag{3.29}$$

$$\overrightarrow{C} = 2\overrightarrow{r}_2 \tag{3.30}$$

where,
$\overrightarrow{a}$ components : are linearly decreased from 2 to 0 over the iterations.
$\overrightarrow{r1}$, $\overrightarrow{r2}$: random vectors in [0, 1].

b. Hunting the prey :

The α wolf usually leads the mission. The β, and δ wolves may occasionally engage with the α wolf .

We assume that the leader (best solution),β, and δ have outstanding knowledge of prospect prey sites, in order to mathematically simulate GW hunting strategy.

As a result, we maintain the initial three most effective solutions acquired and oblige the other search agents (which include the ω) to refresh their positions corresponding to the best search agents' locations.

The formulas of the hunting process are detailed bellow :

$$\overrightarrow{D}\alpha == |\overrightarrow{C}_1 \cdot \overrightarrow{X}\alpha - \overrightarrow{X}|, \overrightarrow{D}\beta == |\overrightarrow{C}_2 \cdot \overrightarrow{X}\beta - \overrightarrow{X}|, \overrightarrow{D}\gamma == |\overrightarrow{C}_3 \cdot \overrightarrow{X}\gamma - \overrightarrow{X}| \tag{3.31}$$

$$\overrightarrow{X}1 = \overrightarrow{X}\alpha - \overrightarrow{A}1 \cdot (\overrightarrow{D}\alpha), \overrightarrow{X}2 = \overrightarrow{X}\beta - \overrightarrow{A}2 \cdot (\overrightarrow{D}\beta), \overrightarrow{X}3 = \overrightarrow{X}\gamma - \overrightarrow{A}3 \cdot (\overrightarrow{D}\gamma) \tag{3.32}$$

$$\overrightarrow{X}(t+1) = \frac{\overrightarrow{X}_1 + \overrightarrow{X}_2 + \overrightarrow{X}_3}{3} \tag{3.33}$$

The suggested hunting approach enables candidate solutions to determine the prey's potential position[89].

c. Exploration and Exploitation in GWO :

The variable C is the principal parameter of GWO to stimulate exploration.

This variable always produces a random value between [0, 2]. It modifies the role of the prey in determining the next location.

When $C > 1$, this contribution is significant because the solution shifts further toward the prey. This parameter generates random values independent of the number of iterations, so exploration is prioritized.

A is a second influencing component that affects exploration. This parameter's value is determined based on $\overrightarrow{a}$, that decreases linearly from 2 to 0.

Because of the random constituents in this parameter, the range for the parameter A varies between [-2 , 2].

When $A > 1$ or $A < -1$, exploration is encouraged, but exploitation is emphasized when $-1 < A < 1$.

To achieve an efficient estimation of the global optimum adopting the stochastic algorithms, a good equilibrium of exploration and exploitation is crucial.

This is accomplished in GWO by decreasing the impact of the A parameter in the formula.

- **GWO Application :**

a. Hyperparameter Tuning

The primary idea behind hyperparameter tuning is to discover the best set of parameter values to enhance the performance of a particular algorithm.

An interesting application of GWO might be to tune the parameters of deep learning algorithms or pre-processing procedures.

b. Feature Selection

Another deep learning-related use of GWO is feature selection.

The method works by creating a list of potential features and iteratively updating these features depending on some performance metric until the desired solution is discovered[88].

Figure 3.12 illustrates the distribution of the use of GWO algorithm by domain of application.

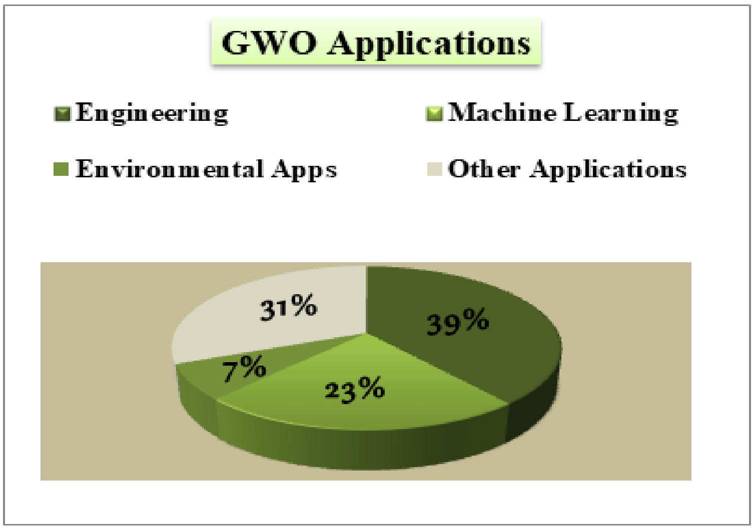


FIGURE 3.12: The use of GWO algorithm by application domain.

This strategy conserves a significant amount of time since the complexity of selecting the best subset of characteristics grows exponentially with each additional variable.

3.2.4 Short-range Wind Speed Prediction: Deep Learning Approach

Traditional statistical procedures are used in wind prediction, which include extra information from weather prediction strategies. These methods have been thoroughly inspected and proven to be reliable in the industry.
The introduction of neural networks in prediction first appeared in the last decade, but regardless of the field fast progress, this new concept is still not extensively maintained by all wind predictors, who prefer employing huge ensembles of approaches using statistical and NWP prediction models instead.
In this section, we will look at the most recent employed DL techniques in the wind speed prediction.

3.2.4.1 Convolutional Neural Network (CNN)

a. Convolution Operation:

Based on definition Convolution is an integral transform of the product of two functions, one of which is reversed and shifted (Eq.3.34).

$$S(t) = (F * G)(t) = \int_{\infty}^{\infty} F(a)G(t-a)da \tag{3.34}$$

Defined as a weighted average of $F(a)$ at time t, where the weighting is supplied by $G(-a)$ with a t shift based on the value of t, the weighting function prioritizes various elements of the input function.

In CNN theory, the function F is defined as the input, the function G is known as the kernel, and the output as the feature map.

While working with data, t will be discretized, and only accept integer values. In such scenario, the discrete convolution is detailed as follows:

$$S(i,j) = (F * G)(i,j) = \sum_m \sum_n F(m,n)G(i-m, j-n) \tag{3.35}$$

When the input is a multidimensional array, the kernel is same, these arrays refer to tensors. Input and kernel components are maintained distinctly. Both functions are considered as zero everywhere outside the limited set of points where the values are maintained.
As a result, the infinite summation may be stated as a summation over a finite number of array members [90].
In this context, convolution is presented as:

$$S(i,j) = (F * G)(i,j) = \sum_m \sum_n F(i-m, j-n)G(m,n) \tag{3.36}$$

b. Pooling :

After executing multiple convolutions same time to generate a collection of linear activations, all linear activations are passed through a nonlinear activation function, such as the rectified linear unit, that referred to the detector step. The pooling function is then employed to substitute the network output at a specific point with a summary statistic of neighboring outputs, to reduce the dimensions of the feature maps.

The assumption is that The existence of a pattern is more essential than its precise position to other patterns. In this scenario, the translation invariance attribute is beneficial. Furthermore, pooling aids in making the representation almost translation invariant.

Pooling across spatial regions results in translation invariance, but pooling over outputs aids the patterns in learning the specific transformations to also become invariant.

Pooling units are often smaller than detector units, considering that pooling uses summary statistics for pooling regions separated by k pixels in the place of 1 pixel. Which implies that the following layer has around k times less inputs to analyze, and it is both computationally and statistically effective. This also reduces the amount of memory required to hold the bottlenecks[91].

c. Activation :

- *Rectified Linear Units (ReLU) :*

The commonly applied and widespread activation function is ReLU. ReLu activation is detailed in Eq.3.37 as:

$$R(x) = max(0, x) \tag{3.37}$$

The ReLU uses real-valued numbers and returns x if x is superior or equivalent to zero, or zero if x is inferior than zero.

At the opposite of the other activation functions, where all the activations are processed, ReLu is considered as cost effective by not tuning a few neurons in the network [92].

d. Loss Function :

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

The loss function evaluates the accuracy of the forecasting outcomes, Where the objective is to reduce the loss function. [93].

In our study, the loss function Mean Squared Error (MSE) is employed to compute the variation between the expected and real results and provided by :

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i) \tag{3.38}$$

Where,

y_i: the real values
$\hat{y}_i$: the resulted values.

3.2.4.2 Recurrent neural network (RNN)

The ideal strategy to deal with sequence to sequence learning applications is Recurrent Neural Networks (RNN).

RNN is considered as a modern form of DL neural networks that can address time series-dependent issues which regular neural networks cannot. It offers a more optimized design for sequence modeling, working on controlling the inputs sequentially, since its internal structure allows for time-related feeding.

In a classic feed forward NN (FFNN) the signal goes in a single direction. Starting from the input layer of the network model, then from the input to the hidden layer, and lastly from the output layer. All layers are totally connected during the signal flow, however nodes in the same layer are not. RNNs vary from feed forward neural networks in that they connect neurons in the recurrent layers, allowing it to recall prior input.

RNNs can successfully evaluate time series and employ the network's memory function to handle data supplied at whatever given point.
RNN is represented mathematically in the upcoming formula:

The layer's output vector is expressed as follows:

$$o_t = gV_{st} \tag{3.39}$$

Where,

o_t : the output at time t.

s_t : the memory at time t.

The loop layer's output vector st is expressed as follows:

$$s_t = f(Ux_t + Ws_{t-1}) \tag{3.40}$$

Where,

x_t : the input at time t.

Because o_t is a function of s_t, it may be expressed as follows:

$$o_t = gV_{st} \tag{3.41}$$

$$g(Vf(Ux_t + Ws_{t-1})) \tag{3.42}$$

$$g(Vf(Ux_t + Wf(Ux_{t-1} + Ws_{t-2}))) \tag{3.43}$$

$$g(Vf(Ux_t + Wf(Ux_{t-1} + Wf(Ux_{t-2} + ...)))) \tag{3.44}$$

Where:

U : the weight from the input to the hidden layer.

V : the weight from the hidden to the output layer.

W : the weight of hidden layer state at the past moment to the one at the current moment of the hidden layer .

f, g : activation function.

The vector that receives the network's output at the present instant is connected to both the current input and to the hidden layer state at the earlier step [94].

This demonstrates that RNN has a memory function and can easily manages time series challenges.

The RNN neuron has a pair of extensions, the Long-Short Term Memory (LSTM) and the Gated Recurrent Unit (GRU), described in the upcoming section:

3.2.4.2.1 Long Short Term Memory Network (LSTM) :

Long-Short Term Memory (LSTM) model is a recent and powerful extension of RNN. The basic RNN model trains the network using back propagation arithmetic.

While working on large sequences, the size of the returned values drops exponentially, causing the slow update of the network weights and the vanishing gradient issue [94].

The process of training the networks usually updates the weights based on the impacts of the latest input signal on the output, with the last signal having more effect on the weight change than the earlier one, this influences the final training outcome .

The fact that training outcomes are heavily influenced by new input data shows that RNNs lack a long memory function.

As a result, Hochreiter et al [95], created the LSTM model to overcome the gradient vanishing issue of RNN and to provide the network with a longer memory function. LSTM basic structure is presented in Fig 3.13.

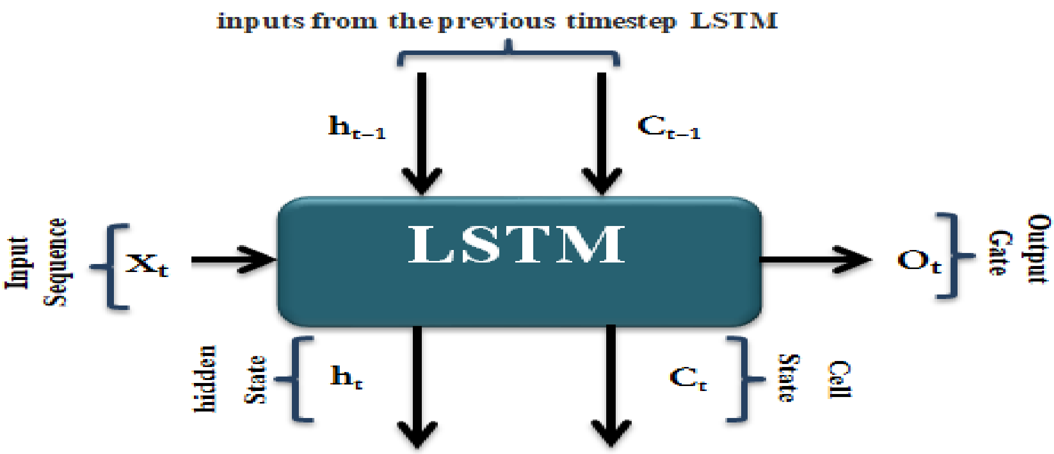


FIGURE 3.13: LSTM Basic architechture.

LSTM is an enhaced version of RNN that employs memory cells at the place of neurons used in the RNN's implicit layer to realize memories of previous data.

The model strategy manages the input, forgetting, and output information of RNNs, that thoughtfully recall and forget the previous information, considerably increasing the learning capacity and application scope of RN networks in time series.

The LSTM adds a state c called unit state, which may retains long-term states, extending it is determined by the time dimension.

The LSTM output comprises two components at the current time: the LSTM output value h_t and the cell state c_t. The LSTM's essential feature is the ability to regulate the long-term state c.

The LSTM employs three control switches:
The first switch in charge of controlling the continued preservation of the long-term state, the second switch controlling the input of the immediate state to the long-term state c, and the third one in charge of controlling whether the long-term state c is used as the current LSTM's output.

Generally, the input, forgetting, and output processes may be used to define the forward transfer process of an LSTM in a single time step [96].

The input gate decides which unit has to be updated (Eq.3.45).

$$i_t = \sigma(W_{xi}x_t + W_{hi}x_{t-1} + W_{ci}x_{t-1} + b_i) \tag{3.45}$$

where,
W : the weight matrix.
b : the offset.

The forget gate f_t defines the past data that must be preserved (Eq.3.46).

$$f_t = \sigma_g(W_f \times x_t + U_f \times h_{t-1} + b_f) \tag{3.46}$$

The state c_t of the network is updated as follows:

$$c'_t = \sigma_c(W_c \times x_t + U_c \times ht - 1 + b_c) \tag{3.47}$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot c'_t \tag{3.48}$$

The output gate o_t is calculated as follows:

$$o_t = \sigma_g(W_o \times x_t + U_o \times h_{t-1} + b_o) \tag{3.49}$$

h_t is obtained as follows:

$$h_t = o_t \cdot \sigma_c(c_t) \tag{3.50}$$

Where σ is the activation function.

With the above notions, it can be assumed that the issue of the long-time reliance in RN networks can be handled using the LSTM mechanism.

3.2.4.2.2 Gated Recurrent Unit Network (GRU) :

The GRU cell is an enhanced version of the LSTM, introduced by Cho et al, in 2014. This network proved its ability in learning long and short dependence, while avoiding slipping into disappearing and bursting gradients.

In contrast to LSTM, GRU has no distinct memory cell and doted with just a pair of gates: the reset gate and the update gate .

The reset gate aids in determining the number of the past hidden state is taken into account in calculating the actual-state candidate, whereas the update gate merges the input and forget gate in the LSTM to define the degree of the past hidden state to be employed to update the actual state[97].

GRU's basic structure is presented in Figure 3.14 bellow :

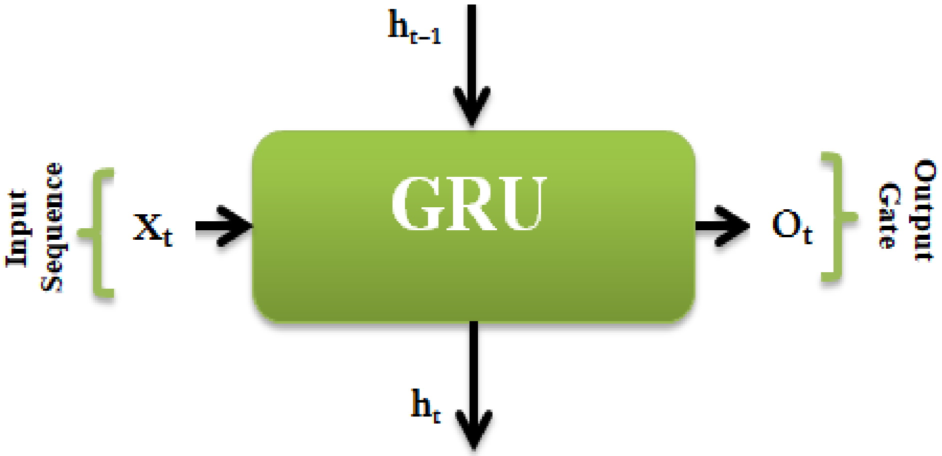


FIGURE 3.14: GRU Basic Structure.

The GRU model is presented by the following formulas:

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h_t} \tag{3.51}$$

$$\tilde{h_t} = g(W_h x_t + U_h(r_t \odot h_{t-1})) + b_h) \tag{3.52}$$

With the update gate and reset gate presented as:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \tag{3.53}$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \tag{3.54}$$

Due to the absence of an output gate, GRU cells are simpler than LSTM.

Although the LSTM has higher learning capabilities, GRU can be trained faster and produces results that are at least as good as those of the LSTM for straightforward issues.

The GRU gates control the data flow into and out of the cell, enabling the cell to retain memories from earlier time periods.

It is unknown which of these two networks is more suited to a given situation, necessitating testing on the particular application and data characteristics.

3.2.4.2.3 Encoder Decoder Architecture (ED) :

RNN are employed in encoder-decoder models, another class of neural networks, to predict outcomes for sequential series that includes time-series data.

The structure of the encoder-decoder network is detailed in Fig. 3.15 below. where y is the model's output and x its input [98] .

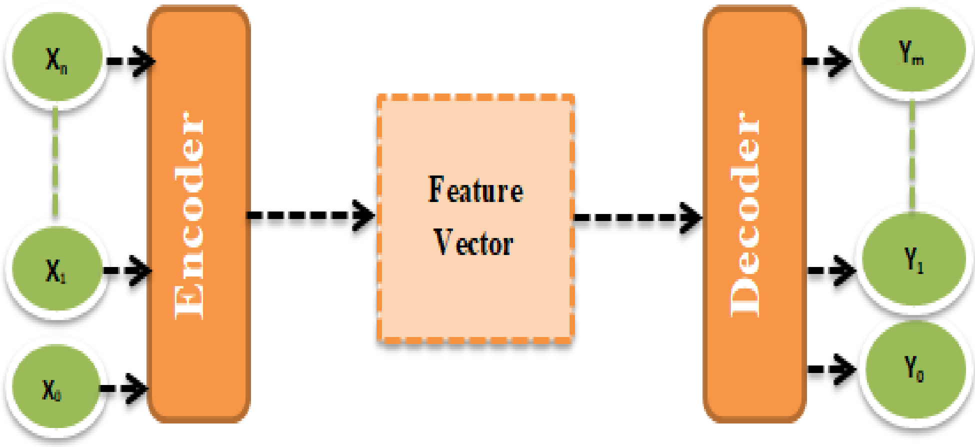


FIGURE 3.15: Encoder-Decoder Architecture.

Three primary elements of the architecture may be identified as follows :

- ***Encoder:*** It takes the components of the input vector one at each time step, learns from the input, and propagates the information to subsequent operations.

- ***Feature Vector:*** This internal and intermediary state is in charge of storing the input's sequential information, which helps the decoder in producing reliable predictions.

- Decoder: The architecture of the decoder component may be an RNN model, that helps in boosting the prediction by decoding the encoder's outcomes in a sequential format again.

Having the LSTM or RNN layer in the encoder decoder architecture, that are originally constructed to work on the sequential data, explains why encoder-decoder models are extremely effective with sequential inputs[98].

We can improve the accuracy of the entire model while working with the sequential data, by simply making the network memorize the sequence with a highly tuned GRU layer [99].

The encoder-decoder approach is then recommended with time series data since it performs well with sequential data.

3.2.4.3 Hybrid Forecasting Models :

Because of the drawbacks of the presented models, the challenging nature of the wind speed data and its highly fluctuated nature, a single model is unlikely to satisfy the accuracy requirement[100].

As a result, the majority of the novel prediction frameworks lays on the notion of hybrid strategies, that lays on combining the pre-processing approaches and DL algorithms [101].

The research on wind speed prediction proved that hybrid strategies outperformed individual forecasting techniques.

The basic concept is that techniques combination balances the shortcomings of one technique with the advantages of the other, although the challenge is to find the best combinations to choose.

3.2.5 Conclusion :

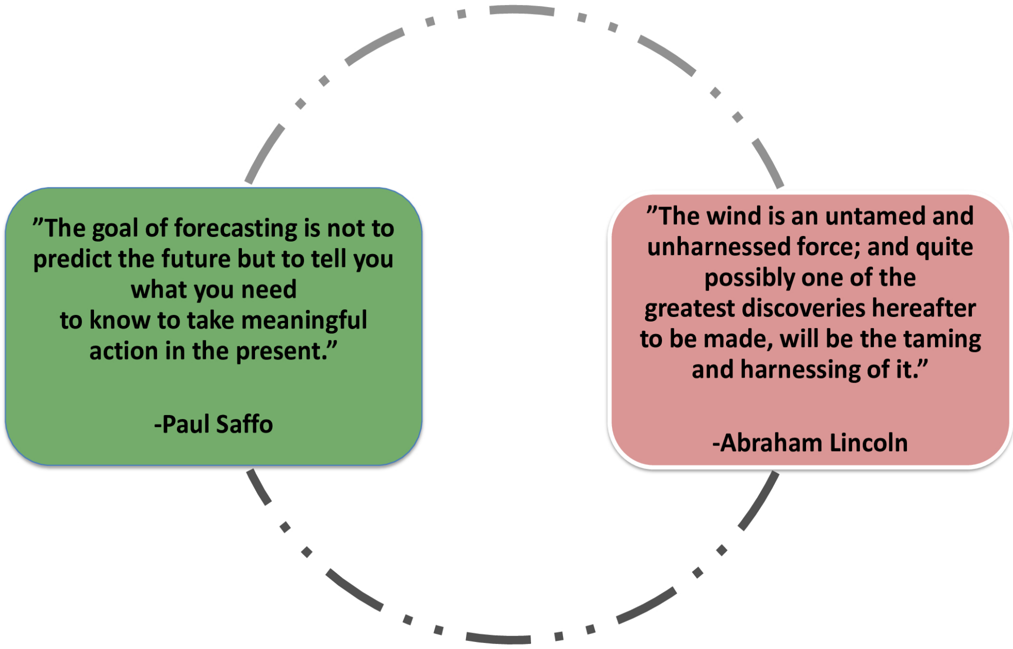
In this chapter we have demonstrated the methods used in the design and development of our proposed forecasting models, we aimed specifically to reach a better comprehension of the wind speed features and the DL forecasting networks, as well as the methods to use for processing them, ranging from data decomposition and denoising to forecasting models construction and improvement.

We have also validated the importance of using hybrid deep learning architectures to reach an efficient wind speed forecast.

To conclude we note that identifying the appropriate combinations and the best hyper-parameters selection for the suggested hybrid architectures, is the key to benefit from prominent time series characteristics, improve forecasting accuracy, and reduce the forecasting costs.

Chapter 4

Contributions



Through the continuous work of researchers around the globe, weather forecasting in general, and precisely wind forecast has known tremendous improvements.
Multiple methods and approaches were invented, upgraded, combined, and improved over the years, with DL as the base of the highly sophisticated strategies used for forecasting.
In this chapter, we introduce the contributions we provided to wind forecasting field, and explain in details the various combinations of models that we proposed through multiple contributions.
These contributions provided a high forecasting accuracy based on numerous experiments.

To improve stability, flexibility, and code readability, the suggested hybrid models are implemented in Python (ver. 3.7.13), offering a variety of libraries and frameworks for machine and deep learning.
For the construction of the neural networks and their functions, we chose Keras (ver. 2.8.0), known as a high-level library.
Keras serves as a wrapper for Tensorflow, a framework specifically designed for machine and deep earning implementations, and that offers fast experimentation, modularity, and scalability, and which also has consistent APIs that minimize the volume of code required for complex functions.
Other Python stacks include : matplotlib.pyplot (version 3.2.2) and numpy (version 1.21.5).....

4.1 Hourly Wind Speed Forecasting Using FFT-Encoder-Decoder-LSTM in South West of Algeria (Adrar)

4.1.1 Introduction

The wind's fluctuated nature makes it a difficult phenomenon to track, and forecasting one of its parameters (wind speed) requires a robust and reliable model.

This first contribution concentrated on wind speed prediction for the purpose of producing wind power[102], that is a critical procedure that demands reliable forecasting findings.

The prediction of wind speed is considered as a highly complex time series issue, where investigations have demonstrated the effectiveness of RN networks, in particular the LSTM model, in making accurate predictions and handling long-term dependencies.

4.1.2 Approach

The presented architecture proposes one hour ahead wind speed prediction model based on FFT Filter (detailed in section 3.2.2.1.2 , p.46 from chapter 3), and Encoder-Decoder-LSTM model (presented in section 3.2.4.2.1 and 3.2.4.2.3 from p.69 and p.73, respectively from chapter 3).

The FFT algorithm was employed for smoothing data process, the Max-Min normalization method (explained 3.2.2.1, p.39 from chapter 3) was used to standardize the signal, and then the Encoder-Decoder-LSTM (ED-LSTM) framework was used to predict wind speed.

Various models were used for comparison purpose to show the effectiveness of the data pre-processing step and the proposed combination. Figure 4.1 illustrates the FFT-ED-LSTM model's whole flowchart.

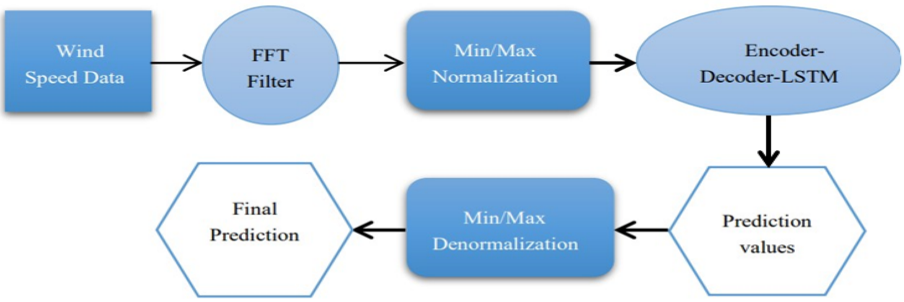


Figure 4.1: The process of the proposed strategy.

4.1.2.1 Dataset and Data Preprocessing

Five years of mean wind speed from Adrar city were employed in this study, provided by Reliable Prognosis Weather for 243 countries of the world website, from 1 January 2014 to 1 January 2019 [103].

Knowing that data was collected every hour, where two-thirds of data was used for training (2014 to 2017) and the rest for test (2017 to 2019).

The FFT denoising effect is illustrated in Figure 4.2 and the process can be summarized as follows:

1. Calculation of FFT from the wind speed original data.

2. Transformation of wind speed signal into frequency range.

3. Elimination of the highest frequency patterns to reach the best smoothing results.

4. The use of inverse FFT to reconstract the original wind speed signal.

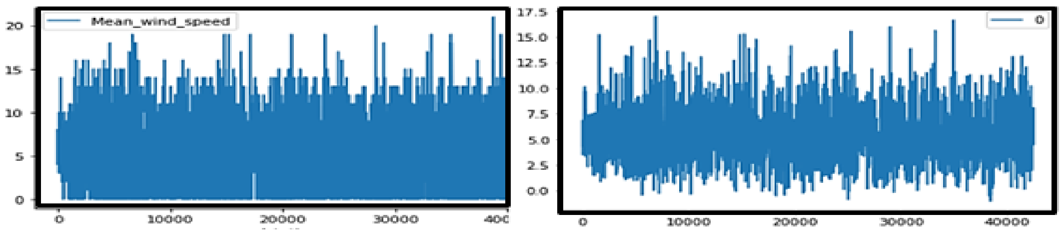


FIGURE 4.2: Wind Speed Data before, and after introducing FFT.

4.1.2.2 Forecasting model : ED-LSTM

The forecasting model combination, employed in this study is shown in Figure 4.3, and made up of two sub-models:

The encoder : "Single LSTM layer" for the input time series sequence's reading and encoding as elements vector.

The decoder : "Single LSTM layer" used to predict each element of the encoded input sequence for a one step ahead.

The LSTM network was used in both the encoding and decoding processes to provide them with complete access to the previous 1 and 3 hours wind speed prediction.

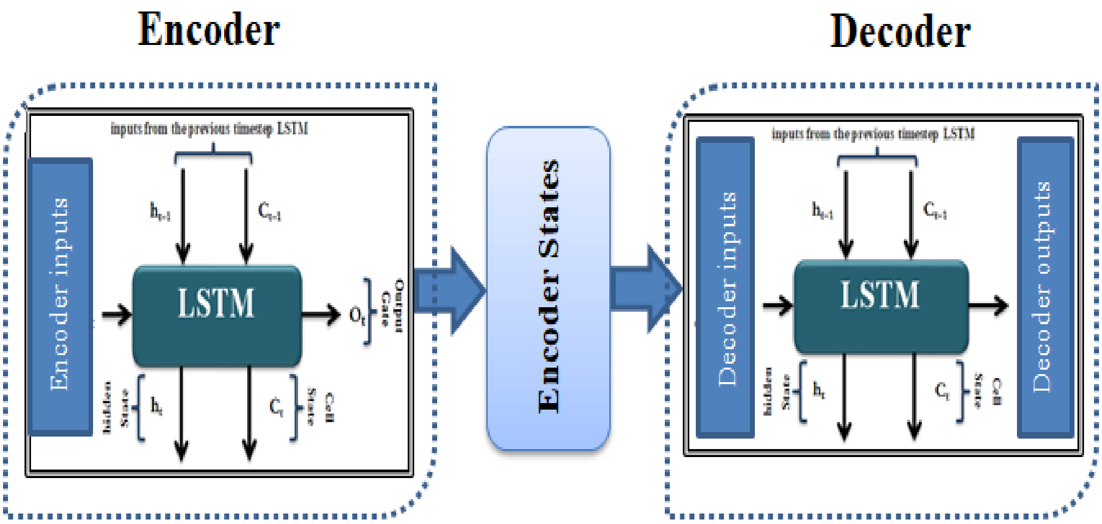


FIGURE 4.3: ED-LSTM architecture.

In this framework, the encoder (single layer LSTM) was given inputs of one and three hours of wind speed data to encode as a vector that captures patterns from the sequence of inputs. For each step in the sequence of outputs, the internal representation of the input sequence for wind speed is recurring several times at the beginning.

After that, the decoder generates the entire sequence, along with all of the input units producing an output value for the wind speed at one and three hours. Every time step in the output series can be interpreted using a fully linked layer.

Finally, the output layer forecasts one step in the output sequence.

- So as to enable the same procedure to recur at every step in the output series, while employing identical weights to create the interpretation, a Time Distributed wrapper is implemented.

4.1.3 Experimentation

4.1.3.1 1-hour ahead wind speed prediction:

Two classifications have been applied in these tests, the first for creating models before applying the FFT algorithm using the MLP, the vanilla LSTM , and the ED-LSTM model, then the following after employing the FFT algorithm combined with each of the previous models.

The first classification is for creating models without applying the Fast Fourier Transform Filter, while the second is for creating models after applying it.

Results are shown in Figures 4.4 and 4.5, and in table 4.1 below:

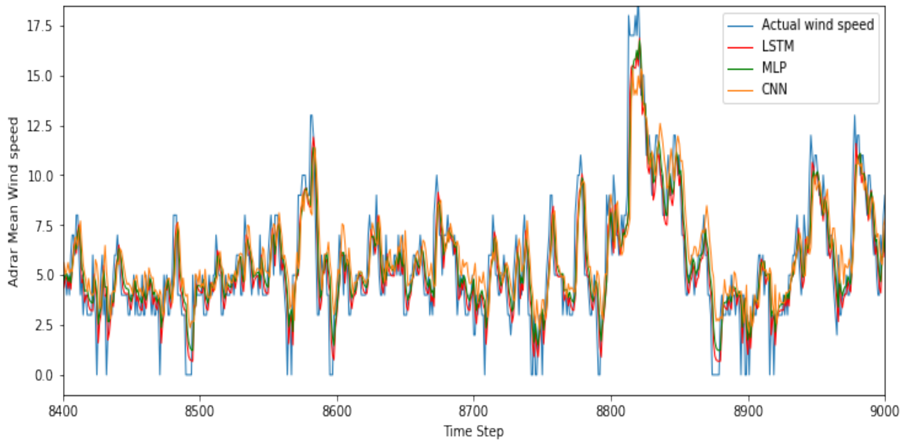


FIGURE 4.4: Single models one-hour ahead prediction.

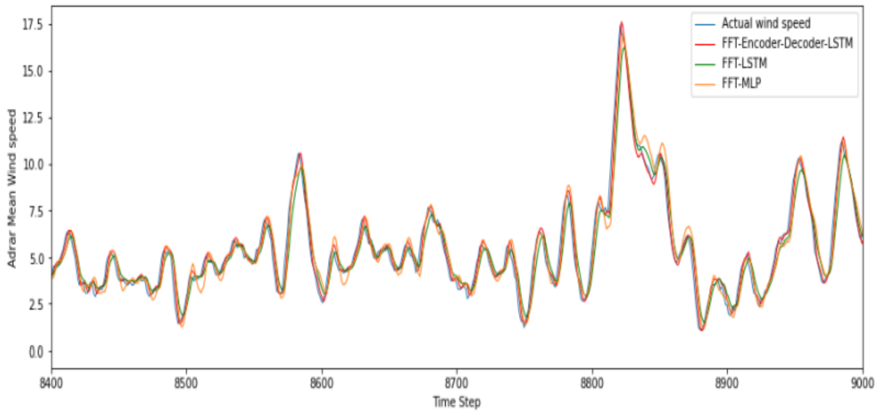


FIGURE 4.5: hybrid models one-hour ahead prediction.

TABLE 4.1: metrics comparison for one-hour ahead prediction.

Algorithms	RMSE (m/s)	MAE (m/s)
MLP	1.981	0.0763
LSTM	1.631	0.0612
Encoder-Decoder-LSTM	**1.432**	**0.0689**
FFT-MLP	**0.194**	**0.009**
FFT-LSTM	**0.043**	**6.3242e-04**
FFT-Encoder-Decoder-LSTM	**0.035**	**6.1734e-04**

4.1.3.2 3-hours ahead wind speed forecasting

For identical tests as in 1-hour ahead forecasting, couple experiments were conducted over a longer time horizon of three hours. Figures 4.6 and 4.7 show the outcomes of different single models.

The forecasting outcomes are reported in Table 4.2 .

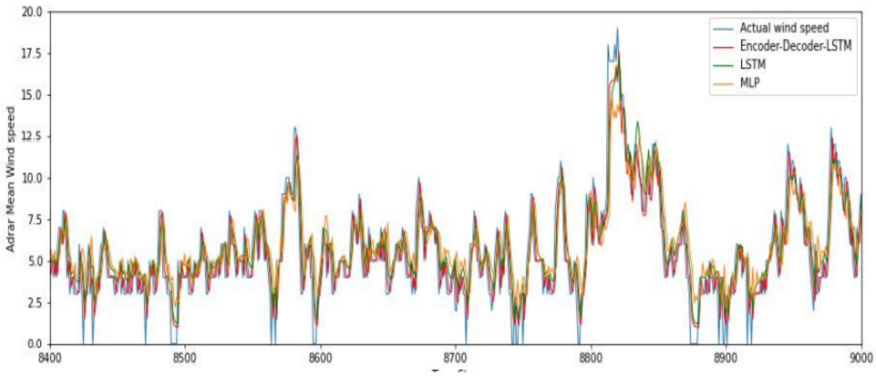


FIGURE 4.6: Single models three-hours ahead prediction.

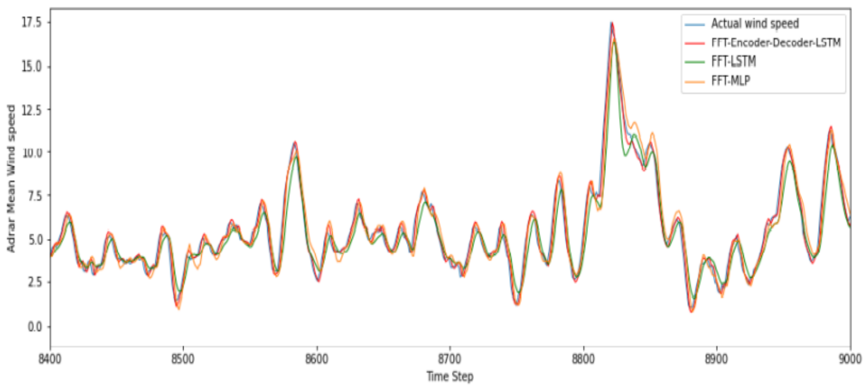


FIGURE 4.7: Hybrid models three-hours ahead prediction.

Table 4.2: Metrics comparison for three-hours ahead prediction.

Algorithms	RMSE (m/s)	MAE (m/s)
MLP	2.313	0.231
LSTM	2.042	0.186
Encoder-Decoder-LSTM	**1.981**	**0.0981**
FFT-MLP	**0.869**	**0.021**
FFT-LSTM	**0.592**	**0.0086**
FFT-Encoder-Decoder-LSTM	**0.487**	**0.0079**

4.1.4 Results Analysis and Conclusion

In our study, we concentrated on wind speed prediction that is crucial for wind energy generation, assuming that, energy produced by the wind is highly related to the cube of its velocity.

We chose the Algerian Sahara's southwest region (Adrar zone) as the study area, having superior wind speed levels and being the first and unique wind park in the entire country.

The outcomes shown in figures and tables above, showed the superior performance of the suggested model, while highlighting the efficacy of the data prep-processing phase in the wind speed prediction process and the effectiveness of the encoder-decoder architecture in providing accurate predictions over different time horizons.

4.2 Wind Speed Forecasting Based on Discrete Wavelet Transform, Moving Average Method and Gated Recurrent Unit

4.2.1 Introduction

Due to the wind's unstable nature, predicting its features stands as a challenging task that necessitates a reliable framework that can learn automatically a sequence of characteristics inside the the wind data.

In this implementation [104], we focused on one hour ahead wind speed prediction employing the GRU model, one of the latest and best evaluated RNN models. This model was chosen because of its simplicity, efficiency, calculation time reduction, and remarkable efficiency.

For smoothing data phase, we made use of two combined methods with high resolution in each of time domains and frequency, being the Discrete Wavelet Transforms (detailed in section 3.2.2.1.2, p.47 from chapter 3), combined with with the Moving Average method (From section 3.2.2.1.2, p.42 from chapter 3).

4.2.2 Approach :

The whole process of the proposed framework is detailed in Fig. 4.8 below:

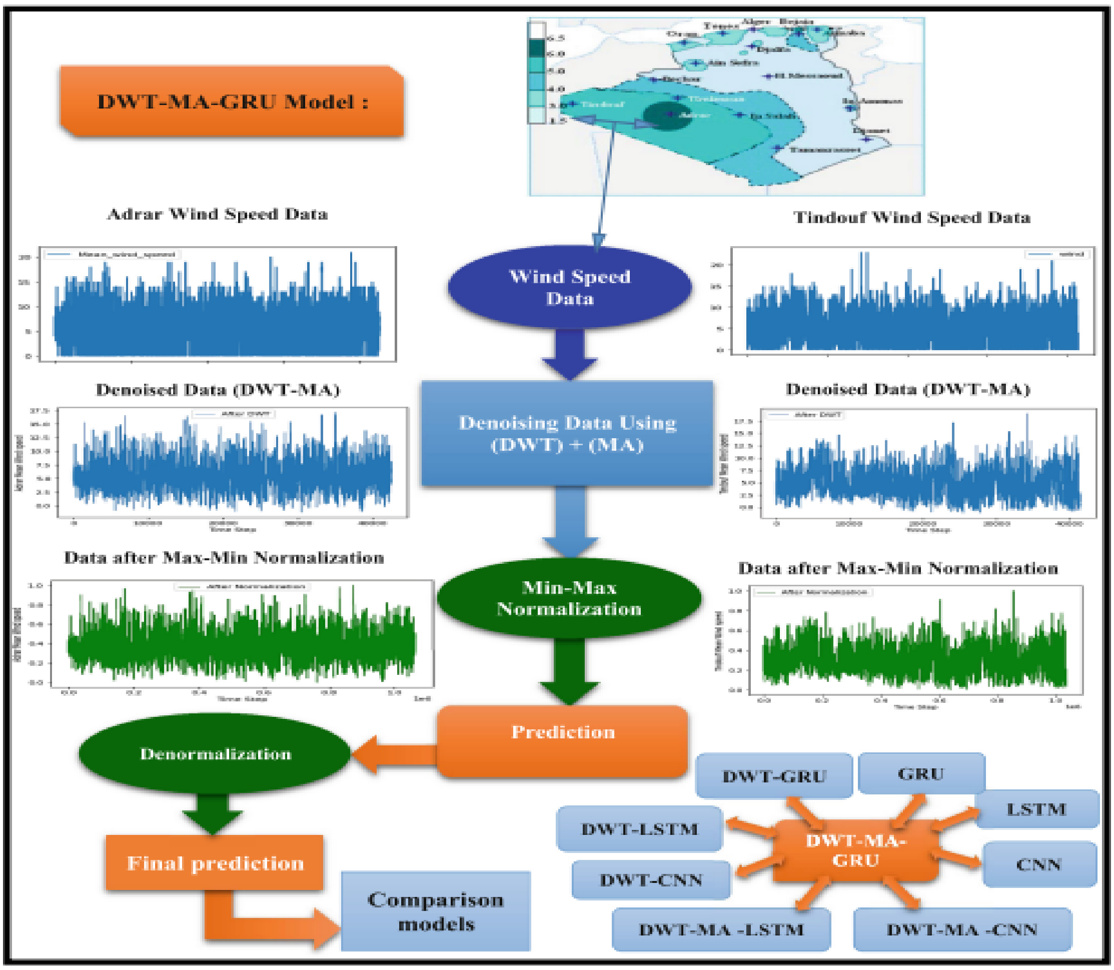


FIGURE 4.8: The structure of the proposed framework.

4.2.2.1 Dataset and Data Preprocessing

Quinquennial of hourly wind speed data were applied in this study, from Adrar and Tindouf cities that have the highest wind speeds over Algeria, to prove the efficiency of the designed architecture over two distinct datasets.

In this study we chose the Daubechies wavelet family, particularly the Daubechies five wavelet (five coefficient filter). The db5 was used up to three levels of decomposition to demonstrate the DWT's power. The procedure for wavelet smoothing strategy employed in this work can be explained as follows:

- Calculating the variance of noise σ starting by the smallest coefficient of wavelet, being d1:

$$\sigma = med(|db5|)/0.6745 \tag{4.1}$$

-The noise threshold limit formula is defined as follows :

$$\gamma = \sigma * sqrt(2 * ln(N)) \tag{4.2}$$

Where,

N : length of the signal

- Soft thresholding is then used to remove noise that is below the estimated threshold limit.

- Finally, the original data is reconstructed applying an inversed wavelet transformation, resulting the denoised signal.

The degree of noise in the lowest frequency bands remained higher than the desired after applying DWT denoising, so the MA technique was used in order to achieve the desired denoised effect. The value at moment (t) is determined by averaging the unprocessed recordings made at the same and previous moment (t) over a tag of nine, as follows :

$$\begin{aligned} \gamma(t) = mean(&obs(t-8), obs(t-7), obs(t-6), obs(t-5), obs(t-4), obs(t-3),\\ &, obs(t-2), obs(t-1), obs(t)). \end{aligned} \tag{4.3}$$

$Y(t)$ presents the outcome of data denoising using MA at moment t, and obs is the record at each moment t.

4.2.2.2 Forecasting model : Gated Recurrent Unit

The proposed forecasting model in this study is composed of a one layer GRU, with the Adam optimizer, MSE, and ReLU, as the network tuning functions.

4.2.3 Experiments

The wind speed in Adrar and Tindouf cities was predicted using three different classifications: before going through preprocessing phase (LSTM, CNN, GRU), after applying DWT technique, after combining DWT and MA method.

4.2.3.1 Adrar Wind Speed Prediction

The outcomes are presented in figures 4.9, 4.10, and table 4.3 below :

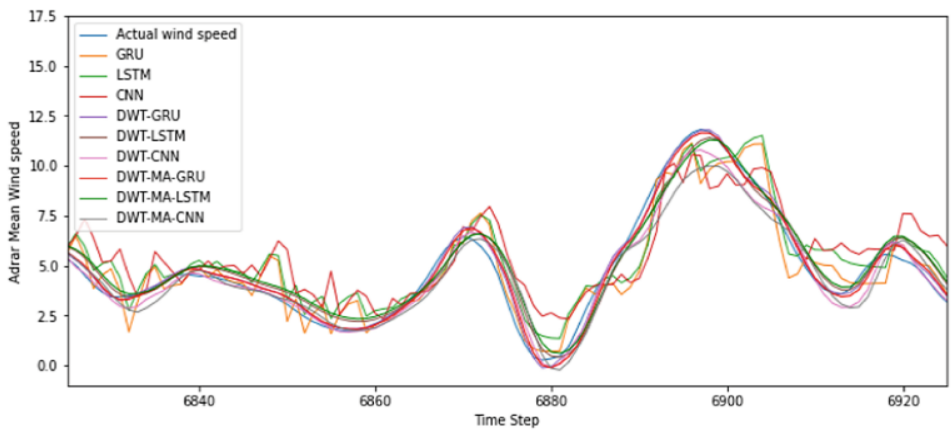


FIGURE 4.9: Forecasting results for Adrar zone.

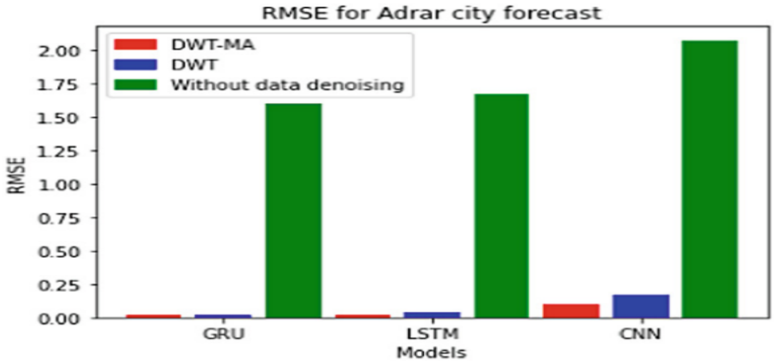


FIGURE 4.10: Metrics outcomes for Adrar zone.

4.2.3.2 Tindouf Wind Speed Prediction

The outcomes are presented in figures 4.11, 4.12, and table 4.3 below :

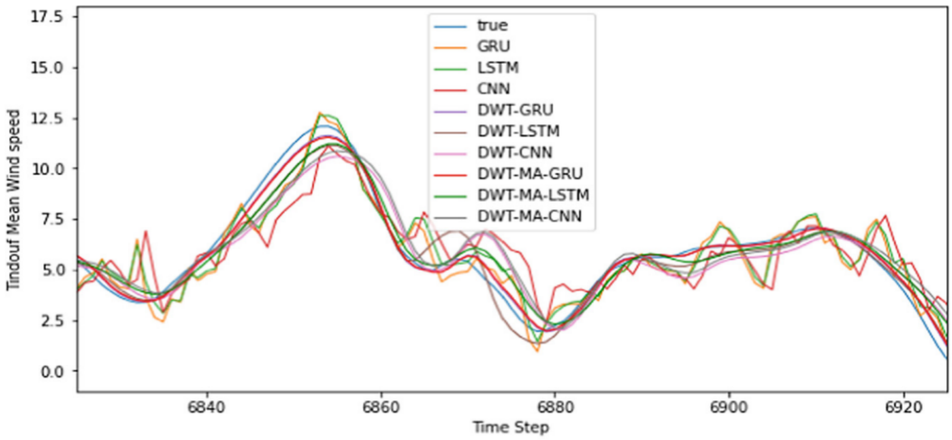


FIGURE 4.11: Forecasting results for Tindouf zone.

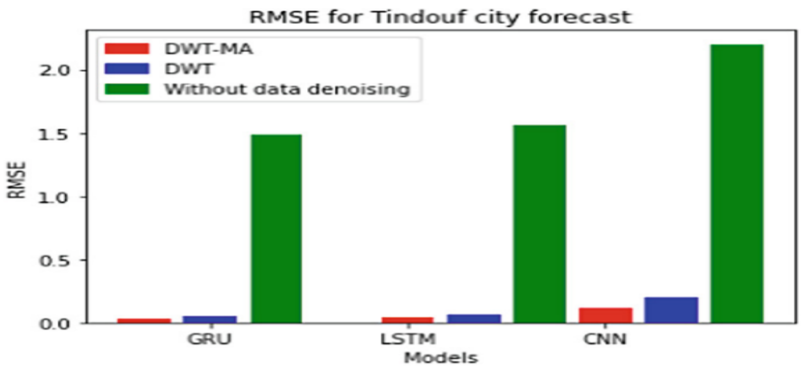


FIGURE 4.12: Metrics outcomes for Tindouf zone.

TABLE 4.3: Metrics outcomes of both cities

Adrar			Tindouf		
Models	**MAE (m/s)**	**MAPE (m/s)**	**Models**	**MAE (m/s)**	**MAPE (m/s)**
CNN	0.0693	22.0132	**CNN**	0.0755	21.0652
LSTM	0.0792	16.0408	**LSTM**	0.0692	14.8603
GRU	**0.071**	**10.9617**	**GRU**	**0.0663**	**11.0167**
DWT-CNN	**0.0080**	**2.8060**	**DWT-CNN**	**0.0080**	**2.6100**
DWT-LSTM	**0.0026**	**0.6923**	**DWT-LSTM**	**0.0026**	**0.7723**
DWT-GRU	**0.0022**	**0.7610**	**DWT-GRU**	**0.0024**	**0.7501**
DWT-MA-CNN	**0.0071**	**2.6763**	**DWT-MA-CNN**	**0.0071**	**2.5773**
DWT-MA-LSTM	**0.0022**	**0.7192**	**DWT-MA-LSTM**	**0.0022**	**0.7309**
DWT-MA-GRU	**0.0020**	**0.6421**	**DWT-MA-GRU**	**0.0018**	**0.5668**

4.2.4 Results Discussion

In this study, a short-range wind speed prediction framework was created by merging the DWT technique, the MA approach, and the GRU network to reach the designed DWT-MA-GRU architecture.

We employed Adrar and Tindouf zone wind speed data, and used the DWT and MA to denoise the datasets before designing the GRU network to forecast the hourly wind speed.

Benchmark models were chosen for comparison. The suggested DWT-MA-GRU provided the most efficient results with the comparision metrics reaching 0.015 (m/s), 0.0018 (m/s), and 0.5668 (m/s), over the two datasets, proving the efficiency of the adopted double denoising approach and the high performance of the GRU model for wind speed prediction.

4.3 Multi-Step Wind Speed Forecasting Based on Hybrid Deep Learning Model and Trailing Moving Average Denoising Technique

4.3.1 Introduction

Finding the precise values of the wind speed is necessary to construct an effective wind power generation architecture, since the wind speed parameter determines how much power the wind farms produce.

In this study [105], we designed a new combined approach that merges the Trailing Moving Average (explained in section 3.2.2.1.2 , p.42 in chapter 3) with the ED-Convolution Neural Network (detailed in section 3.2.4.1 , p.65 in chapter 3) and the LSTM model for multi-step ahead wind speed prediction.

This framework makes use of the CNN model's high capacity for reading through input series, and its ability to learn the most significant characteristics from the feature maps, as well as the LSTM network, which learns features automatically from a sequence within time series.

4.3.2 Approach

The full architecture of the developed framework is expressed in Figure 4.13.

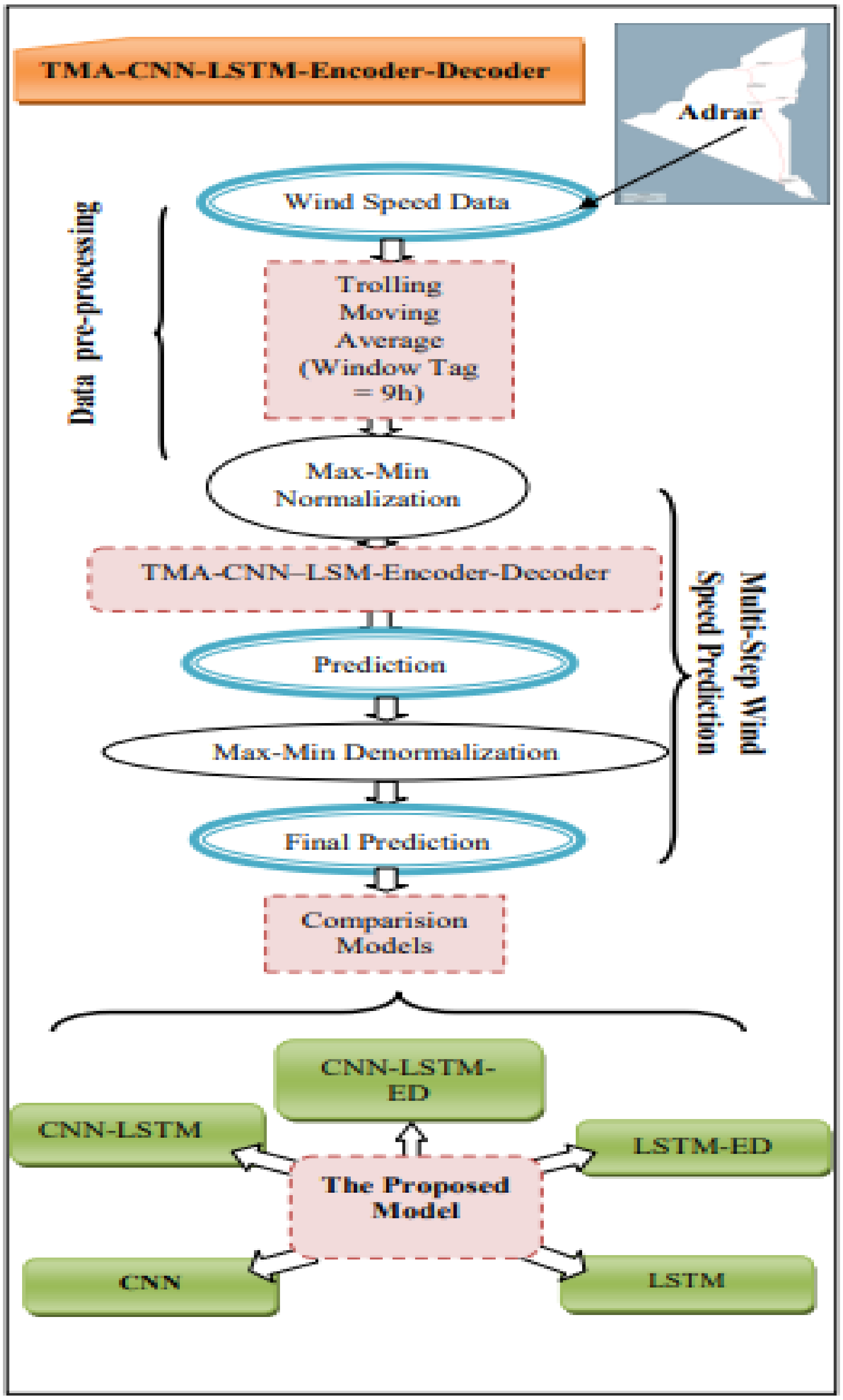


FIGURE 4.13: The structure of TMA-ED-CNN-LSTM framework.

4.3.3 Dataset Data preprocessing

For these experiments, mean wind speed data was gathered from Adrar city from the period between 01/01/2014 and 01 /01/2019. 60% of the wind speed data were set to train the model and the rest 40% for test.

The map from the Algerian Wind Atlas was used as the basis for the site selection[106], indicating great potential for the study.

In this study, we used the TMA approach to smooth the noisy wind speed signal.

Figure 4.14 and 4.15 illustrate the wind recordings with and without passing through the pre-processing phase.

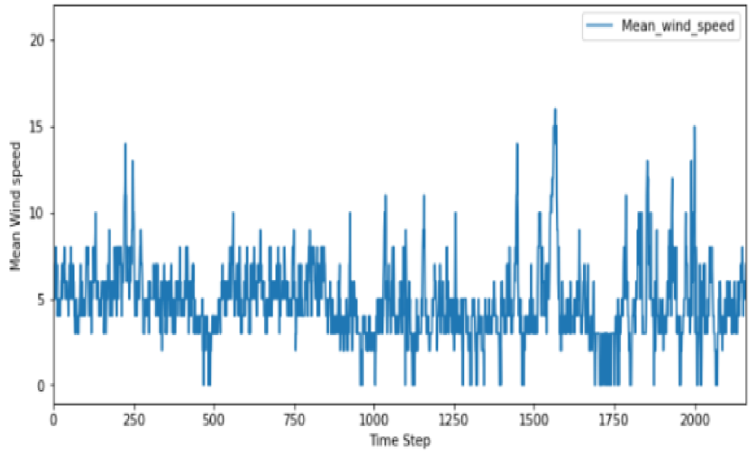


FIGURE 4.14: Original Data.

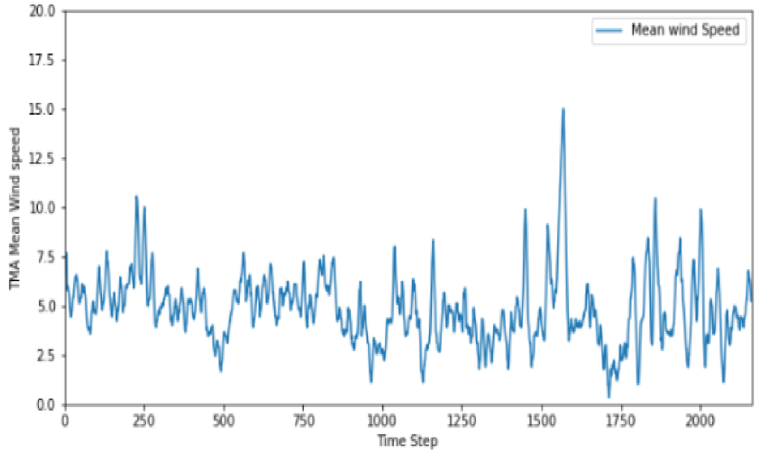


FIGURE 4.15: smoothed Data .

4.3.3.1 Forecasting model : Stacked-ED–CNN-LSTM

In our designed architecture demonstrated in Figure 4.16 :
- The encoder : composed of a Double-layers-CNN network.
- The decoder : consists of Single-Layer-LSTM model.

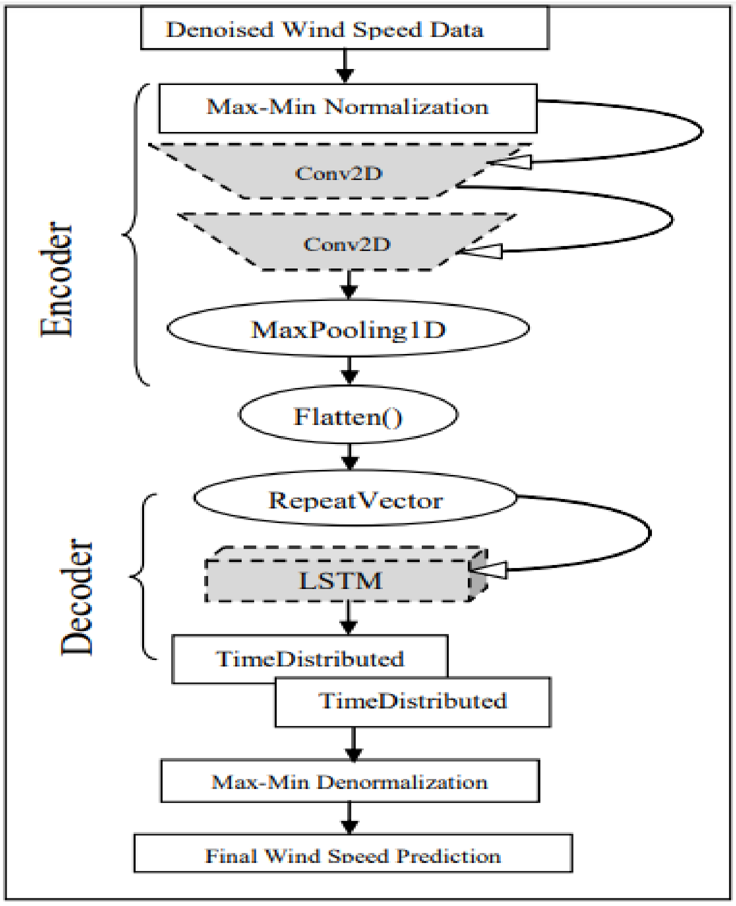


FIGURE 4.16: Encoder-Decoder CNN- LSTM prediction model.

4.3.4 Experimentation

All the forecasting models outcomes from 1-step to 3-steps ahead are shown in Tables 4.4, 4.5, and Figures 4.17, 4.18, 4.19, 4.20, 4.21 and 4.22 below.

TABLE 4.4: The RMSE comparison outcomes.

Algorithms	**RMSE (m/s)**		
	1-step	**2-step**	**3-step**
LSTM	1.713	1.903	2.151
CNN	2.310	2.415	2.913
CNN-LSTM	1.614	1.815	1.873
LSTM-ED	1.541	1.620	1.587
CNN-LSTM-ED	1.270	1.189	1.214
TMA-CNN-LSTM-ED	**0.384**	**0.327**	**0.231**

TABLE 4.5: The MAE comparison outcomes.

Algorithms	MAE (m/s)		
	1-step	2-step	3-step
LSTM	0.066	0.087	0.112
CNN	0.369	0.514	0.702
CNN-LSTM	0.051	0.024	0.076
LSTM-ED	0.032	0.057	0.068
CNN-LSTM-ED	0.027	0.032	0.028
TMA-CNN-LSTM-ED	**0.009**	**0.014**	**0.007**

4.3.4.1 One-step-ahead prediction results

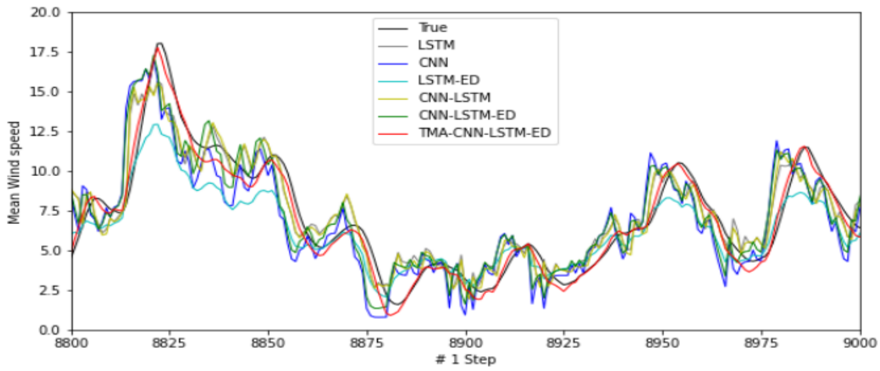


FIGURE 4.17: One-step forecasting outcomes.

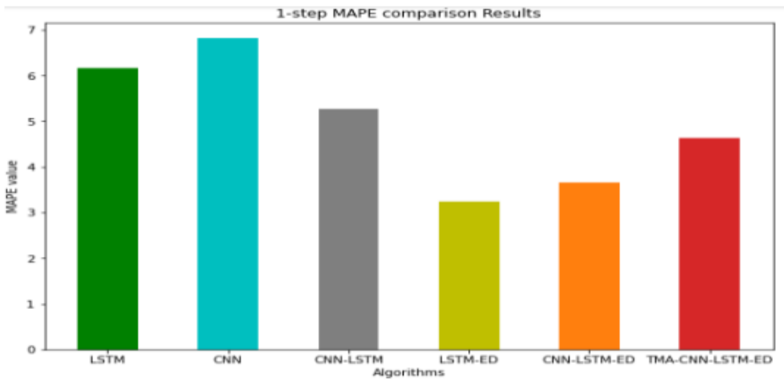


FIGURE 4.18: One-step MAPE forecasting outcomes.

4.3.4.2 Two-steps-ahead forecasting outcomes

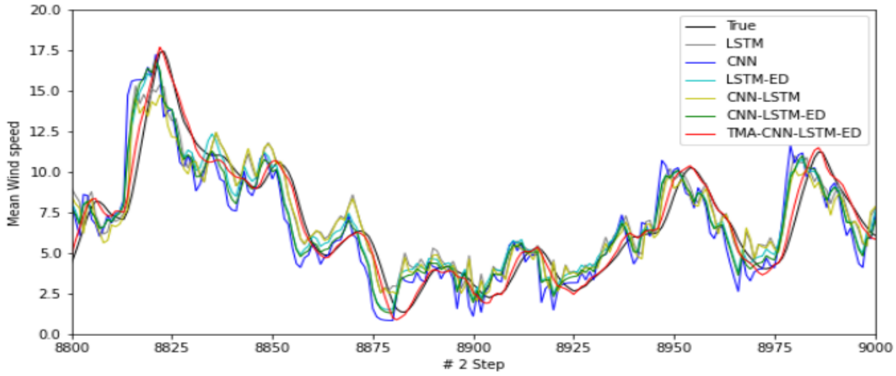


FIGURE 4.19: Two-steps forecasting outcomes.

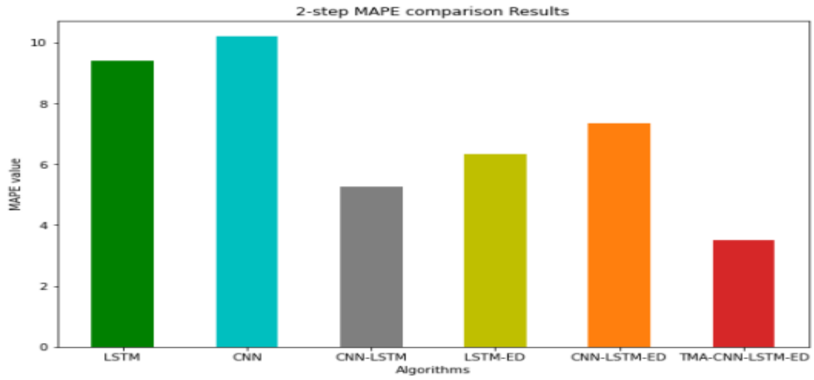


FIGURE 4.20: Two-steps MAPE forecasting outcomes.

4.3.4.3 Three-steps-ahead forecasting outcomes

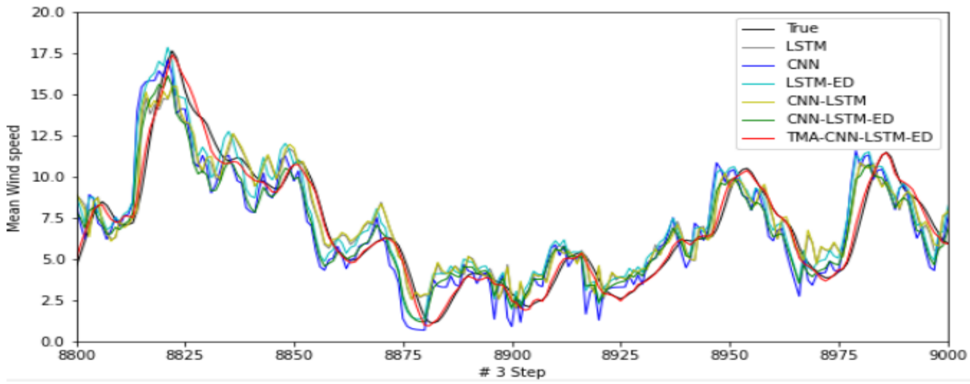


FIGURE 4.21: Three-steps forecasting outcomes.

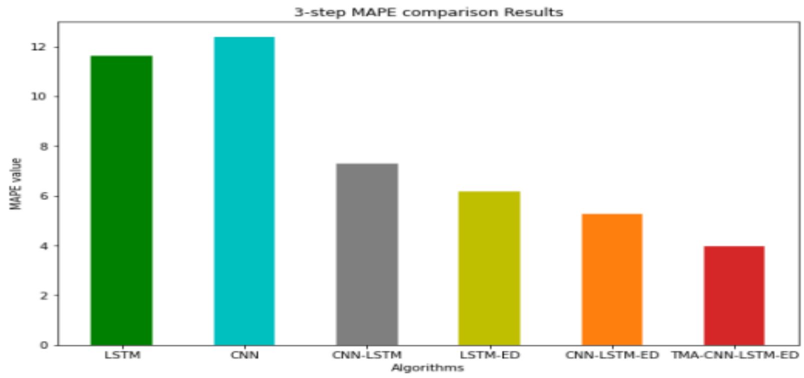


FIGURE 4.22: Threes-step MAPE forecasting outcomes.

4.3.5 Discussion and Conclusion

For the purpose of obtaining the maximum wind speeds in the southwest of Algeria, a hybrid short-range wind speed prediction combination was used in this study.

To prove the superiority of the designed framework, a couple of deep learning models were chosen as benchmark models.

The TMA smoothing technique demonstrated a significant improvement in the prediction accuracy.

The CNN-LSTM-ED combination can extract and predict more features from wind speed data, confirming the performance of the encoder-decoder structure.

The experimental outcomes demonstrated that the suggested framework improved the prediction results remarkably compared to the alternative models.

4.4 Hybrid intelligent framework for one-day ahead wind speed forecasting

4.4.1 Introduction

The advancement of alternative energies technologies depends greatly on the nations that create, produce, and export them.

In this implementation we worked on developing a standard framework that offers the possibility of surpassing the hyper-parameter selection constraint, that is demanding in computation and time, in order to obtain a standard framework capable of forecasting different datasets without the need of adaptation to the different data characteristics [107].

Depending on the Adaptive GWO (explained in section 3.2.3.3, p.60 in chapter 3), Singular Spectrum Analysis (SSA) (detailed in section 3.2.2.1.2 p.56 in chapter 3), with the suggested ED-CNN-GRU, a new hybrid architecture is designed for one day ahead wind speed prediction.

The AGWO-SSA is designed in order to separate the original wind speed series into its patterns and detailed features, to facilitate the oscillate noise removal from the original collected wind speed, the smoothed data is then introduced to the optimized ED-CNN-GRU model developed for 24-hours ahead reliable wind speed forecasting.

Where the AGWO is employed in order to achieve the best parameters selection for the length of the SSA window with the number of layers of the ED-CNNGRU model.

4.4.2 Approach

4.4.3 Dataset and data preprocessing

Approximately ten to twelve-meters high beyond the surface of the ground, mean wind speed data for eight years were gathered from three separate stations, situated in the southwestern zone of Algeria, from September 2012 to December 2020.

The mean wind speed was recorded hourly, where two-thirds of the information was set aside to train the model and the rest for validating and testing the outcomes.

The sites were picked based on the Atlas wind map of the Algerian region, where the favorable wind speeds conditions are available, to test the proposed framework on three various dataset with multiple features.

Table 4.6 and Figure 4.23 provide descriptions of the datasets in detail.

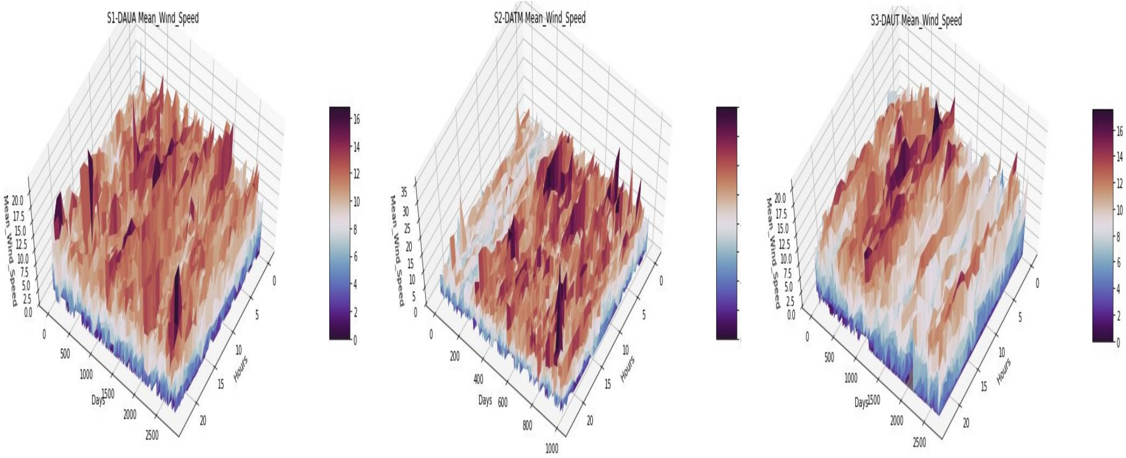


FIGURE 4.23: Stations's Datasets.

TABLE 4.6: Stations's details

Stations	Station Location	Nbr of obs	Mean val	Max val
S1-DAUA	280 m	67325	5.4(m/s)	22(m/s)
S2-DATM	397 m	24013	5.8(m/s)	36(m/s)
S3-DAUT	312 m	64666	4.5(m/s)	21(m/s)

Figure 4.24 bellow, illustrates the full AGWO-SSA-ED-CNNGRU model process.

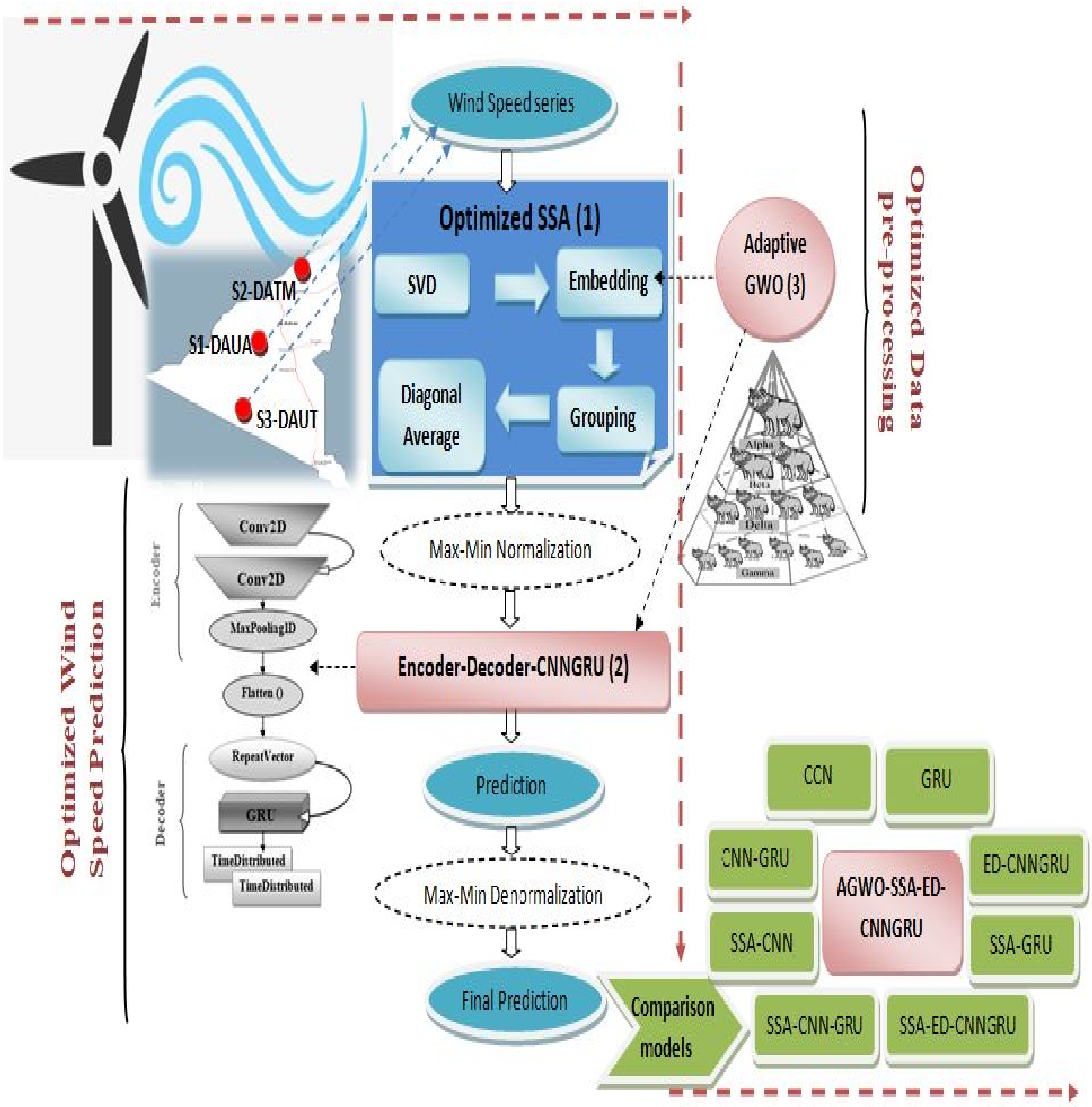


FIGURE 4.24: The architecture of the proposed hybrid strategy.

The hybrid framework architecture can be described as follows :

(1) To get over the reliance on the data variation problem and capture the features and patterns in the original datasets, the simplified structure of the SSA preprocessing method combined with the AWO is used.

In Figure 4.25 bellow, the AGWO-SSA denoising process is illustrated :

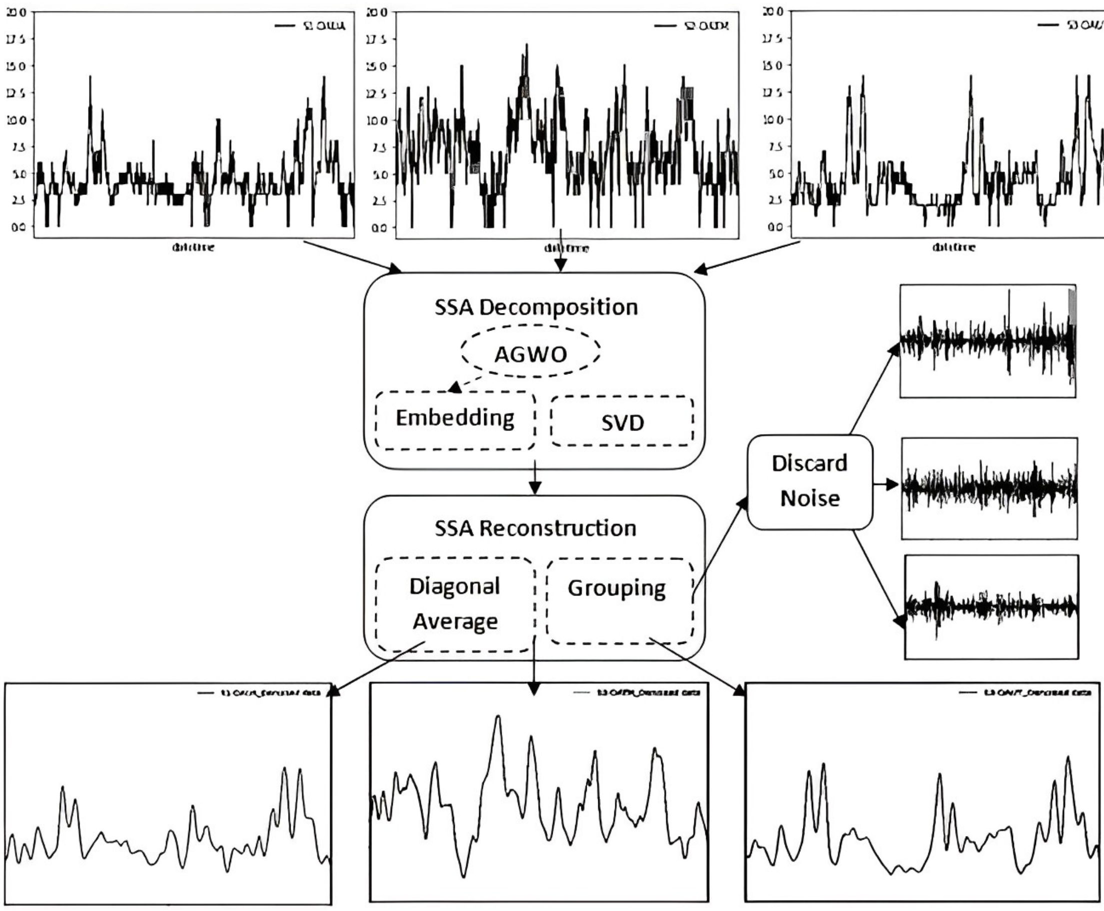


FIGURE 4.25: The AGWO-SSA denoising architecture.

(2) The improved ED-CNN-GRU framework is created specifically to manage the fluctuating nature of the original wind datasets .

The encoder: is a Double-Layers-Stacked-CNN model, that was employed on the original wind input datasets to extract the most prominent features.

In order to read through the input series and display the outcomes as features map, the first CNN layer was created.

The max pooling layer is then designed for the features map simplifying process, after using a second CNN layer to perform an identical treatment on the features map resulted from the previous one, to highlight the significant characteristics.

These extracted flattened feature maps can then be used as inputs for the decoder model, after being simplified.

The decoder: is a GRU network with competitive performance and fast calculation time that can automatically learn characteristics from a sequence inside the time series.

Figure 4.26 explains the proposed model :

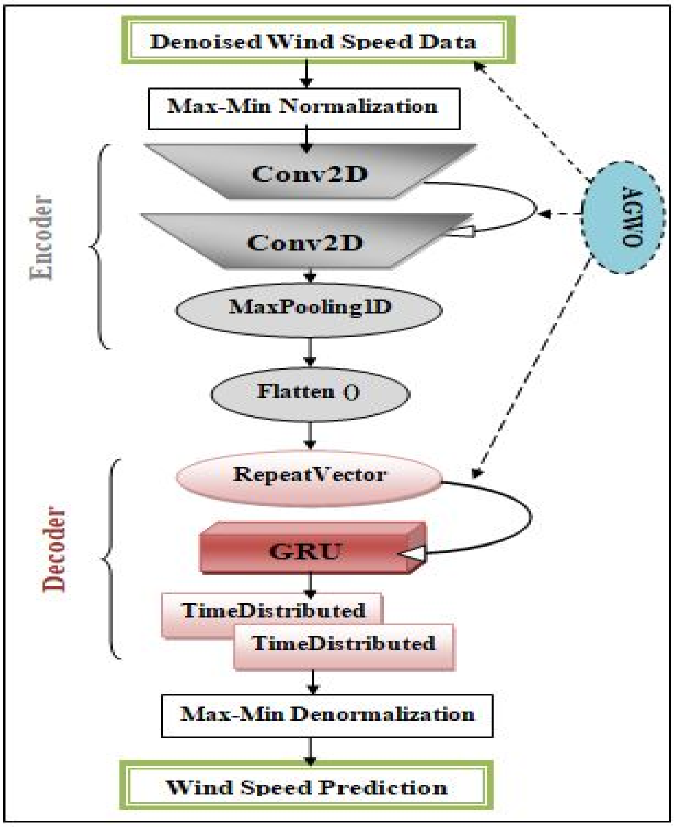


FIGURE 4.26: The hybrid DL forecasting model.

(3) For the hybrid model optimization, meta heuristics were recalled, in particular the Adaptive GWO approach was applied for this study, offering an intelligent parameter tuning for the length of the SSA window and the number of layers for the ED-CNN-GRU prediction model.

Bellow the full pseudo-code of the Adaptive optimizer :

Algorithm 1: Adaptive GWO for SSA and Encoder-Decoder CNNGRU Layers optimization

```
InitWolvesPack();
// each wolf is a combination of a random selected values for both parameters (Window Length
   of SSA and the ED-CNNGRU number of Layers)
MakePrediction();
// Full wind speed values forecasting

for every wolf in the pack do
    calculateFitnessForWolf(); // calculation of the fitness for each wolf of the initial pack :
       The fitness is a combination of 3 values :  RMSE,MAE and MAPE we got after making a
       full prediction of the wind speed
end
SelectAlpha,Beta,DeltaWolves();
// The Alpha,Beta,Delta wolves will have the best RMSE,MAE and MAPE values respectively

while stop criteria is not reached do
    InitializeControlParameterA();
    // Initialization of eter A that controls the exploration in the search for the best
       combination of parameters
    for every wolf except alpha beta delta do
        // First Parameter :  SSA Window Length
        X1p1 = alphawolfP1 - A*(alphawolfP1 – wolfP1) ;
        X2p1 = betawolfP1 - A*(betawolfP1 – wolfP1) ;
        X3p1 = deltawolfP1 - A*(gammawolfP1 – wolfP1);
        // Second eter :  CNN Number of layers
        X1p2 = alphawolfP2 - A*(alphawolfP2 – wolfP2) ;
        X2p2 = betawolfP2 - A*(betawolfP2 – wolfP2);
        X3p2 = deltawolfP2 - A*(gammawolfP2 – wolfP2);
        // Third eter :  GRU Number of layers
        X1p3 = alphawolfP3 - A*(alphawolfP3 – wolfP3) ;
        X2p3 = betawolfP3 - A*(betawolfP3 – wolfP3);
        X3p3 = deltawolfP3 - A*(gammawolfP3 – wolfP3);
        WolfP1 = (X1p1 + X2p1 + X3p1) // 3;
        WolfP2 = (X1p2 + X2p2 + X3p2) // 3;
        WolfP3 = (X1p3 + X2p3 + X3p3) // 3;
        // Update each wolf of the pack depending on the position (values of parameters) of
           Alpha, Beta and Delta wolves except the 3 last mentioned ones
        if wolf parameters get pass boundaries then
            WolfInitizlizationWithRandomParameteres();
            // Initialize current wolf with random values inside the boundaries of each eter
               to keep the search flow in the right direction.
            ;
        else
        end
    end
    MakePrediction();
    for every wolf in the pack do
        calculateFitnessForWolf();
    end
    SelectAlpha,Beta,Delta Wolves();
    // re-define the pack of Alpha, Beta and Delta wolves positions depending on the new
       fitness values calculated in the previous step
end
BestWindowLengthAndLayersNB();
// We select the best combination of ED-CNNGRU layers and SSA Window Length while putting in
   consideration the 3 last saved Alpha, Beta and Delta wolves
```

4.4.4 Experimentation

4.4.4.1 Hyper-parameters tuning

To identify the optimal correlation between the designed ED-CNN-GRU's layers number and the optimized SSA Window-Length, we used the GW optimizer, known for it's stability and reliability among the other commonly used meta-heuristic algorithms, then modified it for our architectural needs.

The framework was trained over 50 iterations.

In order to prevent the randomly initializing impact, the pack of wolves is changed after each iteration.

Based on the GWO principles, the best results are then taken into consideration following the wind speed prediction.

The three most employed metrics : MAPE, MAE, and RMSE are used for the performance validation of our suggested framework, proving the superior results it achieved among the benchmark models.

The following plots of Figure 4.27 and table 4.7 display a sample from the fifty successive iterations that the proposed framework launched to achieve the optimal combination, while also displaying the updates in the triple wolves packs at the end of each iteration.

The following parameters were the best combination of the wolves packs choosed by the adapted optimizer:
The length of SSA Window : 9 .
The number of CNN Layers : 2. The number of GRU Layers : 1.

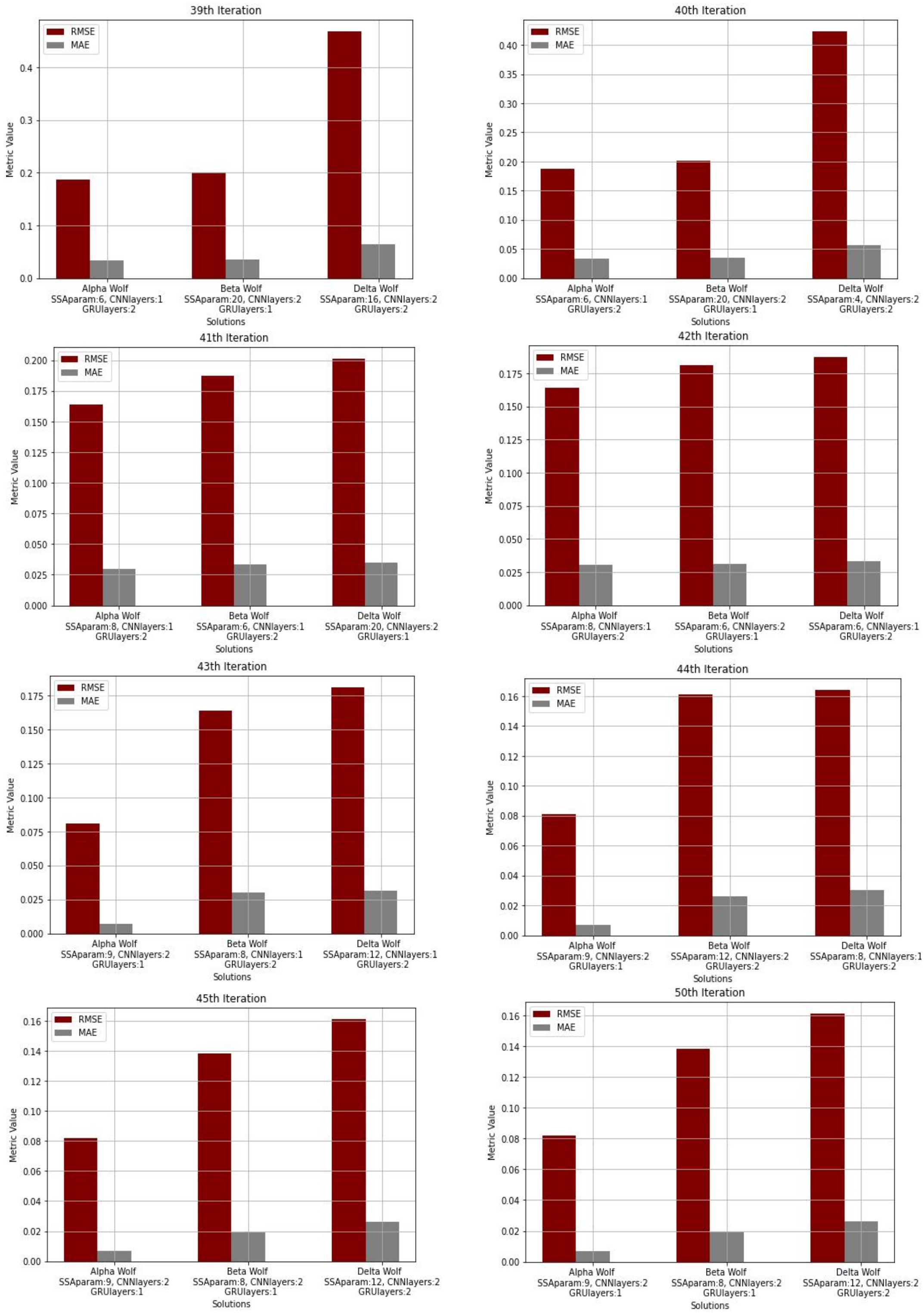


FIGURE 4.27: A sample of the proposed model last iterations.

TABLE 4.7: Best combination tuning

Iteration	Update alpha,beta,delta Wolves ?	Alpha Wolf	RMSE MAE MAPE	Beta Wolf	RMSE MAE MAPE	Delta Wolf	RMSE MAE MAPE
		SSA : 6	- 0.187	SSA : 20	- 0.201	SSA : 16	- 0.468
39th	Yes	CNN : 1	- 0.0332	CNN : 2	- 0.035	CNN : 2	- 0.065
		GRU : 2	- 4.043	GRU : 1	- 4.210	GRU : 2	- 4.391
		SSA : 6	- 0.187	SSA : 20	- 0.201	SSA : 4	- 0.423
40th	Yes	CNN : 1	- 0.0332	CNN : 2	- 0.035	CNN : 2	- 0.057
		GRU : 2	- 4.043	GRU : 1	- 4.210	GRU : 2	- 4.390
		SSA : 8	- 0.164	SSA : 6	- 0.187	SSA : 20	- 0.201
41th	Yes	CNN : 1	- 0.0301	CNN : 1	- 0.0332	CNN : 2	- 0.035
		GRU : 2	- 3.845	GRU : 2	- 4.043	GRU : 1	- 4.210
		SSA : 8	- 0.164	SSA : 6	- 0.181	SSA : 6	- 0.187
42th	Yes	CNN : 1	- 0.0301	CNN : 2	- 0.0312	CNN : 1	- 0.0332
		GRU : 2	- 3.845	GRU : 1	- 3.903	GRU : 2	- 4.043
		SSA : 9	- 0.081	SSA : 8	- 0.164	SSA : 12	- 0.181
43th	Yes	CNN : 2	- 0.007	CNN : 1	- 0.0301	CNN : 1	- 0.0312
		GRU : 1	- 2.481	GRU : 2	- 3.845	GRU : 2	- 3.903
		SSA : 9	- 0.081	SSA : 12	- 0.161	SSA : 8	- 0.164
44th	Yes	CNN : 2	0.007	CNN : 2	- 0.0261	CNN : 1	- 0.0301
		GRU : 1	- 2.481	GRU : 2	- 3.143	GRU : 2	- 3.845
		SSA : 9	- 0.081	SSA : 8	- 0.138	**SSA : 12**	- 0.161
45th	Yes	**CNN : 2**	- 0.007	CNN : 2	- 0.019	**CNN : 2**	- 0.0261
		GRU : 1	- 2.481	GRU : 1	- 2.814	**GRU : 2**	- 3.143
		SSA : 9	- 0.081	SSA : 8	- 0.138	**SSA : 12**	- 0.161
50th	No	**CNN : 2**	- 0.007	CNN : 2	- 0.019	**CNN : 2**	- 0.0261
		GRU : 1	- 2.481	GRU : 1	- 2.814	**GRU : 2**	- 3.143

4.4.4.2 The developed Strategy Evaluation

The following plots (4.28, 4.29 and 4.30) contain the outcomes of one day ahead prediction in 3 stations with various characteristics using the proposed framework.

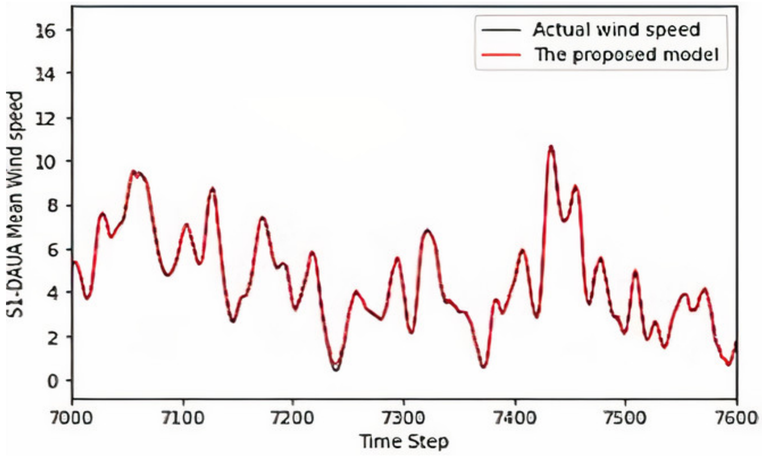


FIGURE 4.28: The first station's forecast.

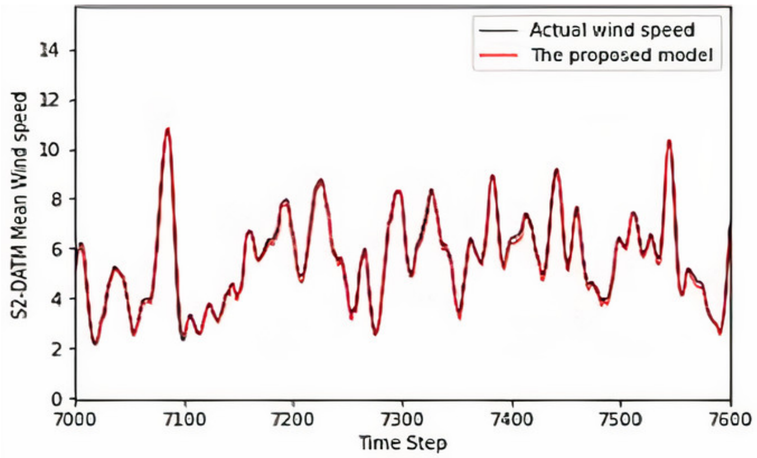


FIGURE 4.29: The second station's forecast.

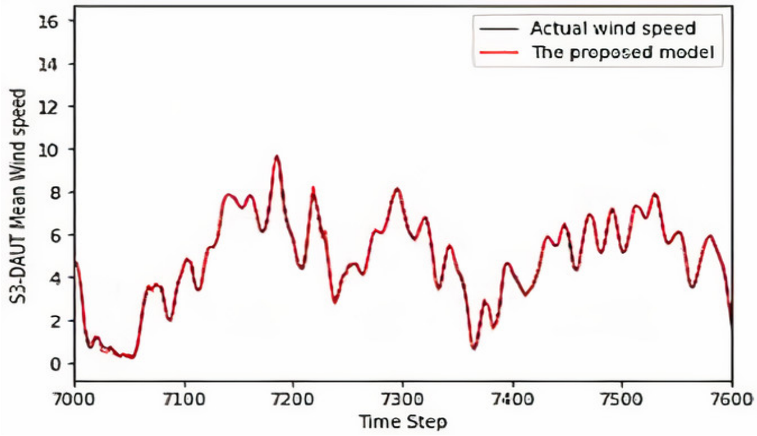


FIGURE 4.30: The third station's forecast.

4.4.4.3 Single Models Forecasting Outcomes

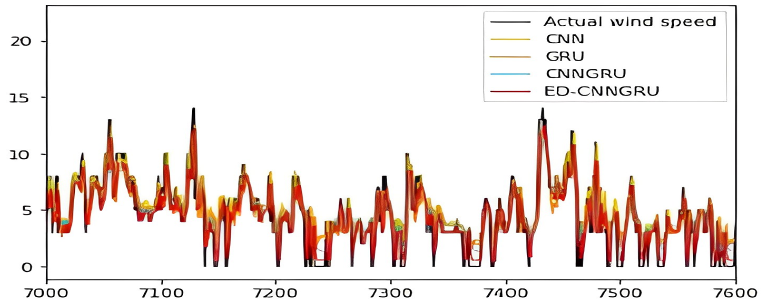


FIGURE 4.31: The first station's comparison results.

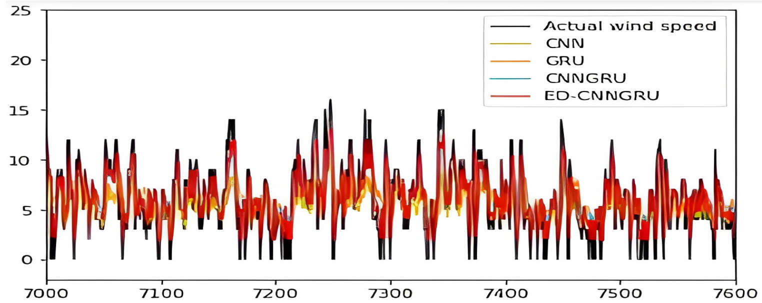


FIGURE 4.32: The second station's comparison results.

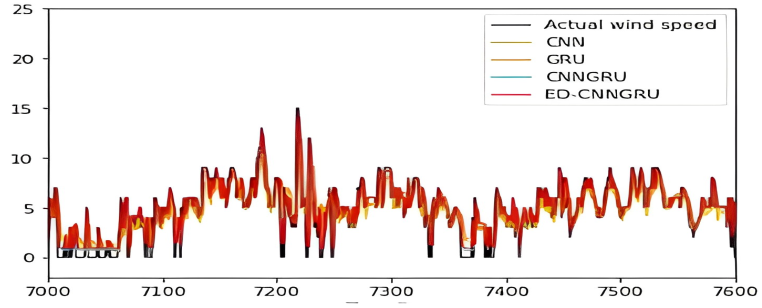


FIGURE 4.33: The third station's comparison results.

TABLE 4.8: Metrics Comparison Results

	Stations								
	S1 DAUA			S2 DATM			S3 DAUT		
Models	**RMSE**	**MAE**	**MAPE**	**RMSE**	**MAE**	**MAPE**	**RMSE**	**MAE**	**MAPE**
CNN	2.8736	0.4867	15.5741	3.7016	0.7174	17.6415	2.93517	0.3742	15.8257
GRU	1.8132	0.1912	13.5471	2.3842	0.2184	14.4261	1.8646	0.2972	13.7514
CNN-GRU	0.8147	0.0731	11.0714	0.9145	0.0814	11.9415	0.8724	0.0725	11.3164
ED-CNNGRU	0.5716	0.0521	9.7431	0.7224	0.0721	10.6451	0.5127	0.0501	9.2571

4.4.4.4 Hybrid models Experiments Results

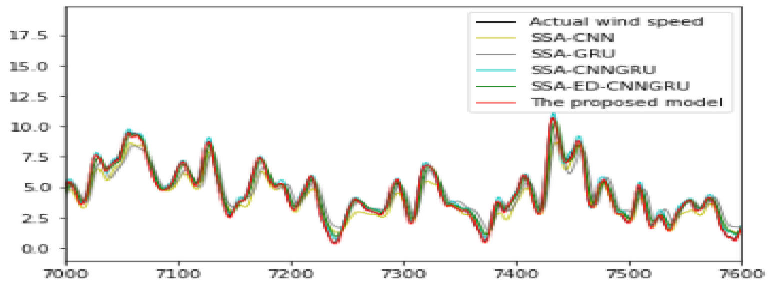


FIGURE 4.34: The first station's forecasting outcomes.

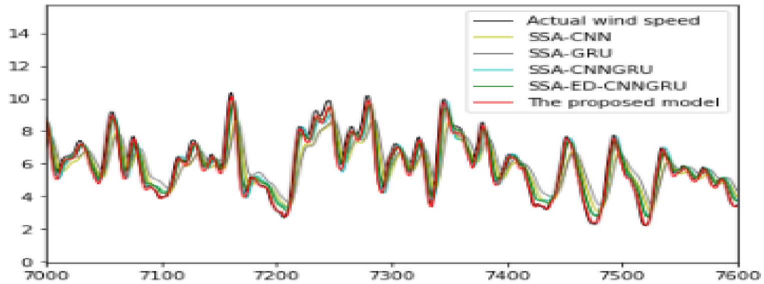


FIGURE 4.35: The second station's forecasting outcomes.

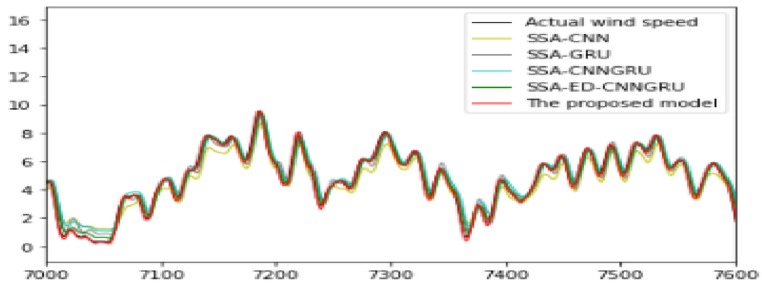


FIGURE 4.36: The third station's forecasting outcomes.

TABLE 4.9: Metrics Comparison for Hybrid Models

	Stations								
	S1 DAUA			S2 DATM			S3 DAUT		
Models	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
SSA-CNN	0.8381	0.1367	11.5741	1.0257	0.3174	13.6415	0.9351	0.1742	12.8257
SSA-GRU	0.5132	0.0712	10.5471	0.7842	0.0984	11.4261	0.6376	0.0872	10.7514
SSA-CNN-GRU	0.2037	0.0614	7.0714	0.3871	0.0871	9.1957	0.2836	0.0589	8.3641
SSA-ED-CNNGRU	0.1603	0.0314	6.2431	0.2224	0.0612	7.5641	0.1927	0.0401	6.7531
The proposed model	**0.0461**	**0.0081**	**4.2431**	**0.0824**	**0.0321**	**5.6491**	**0.0607**	**0.0093**	**4.6571**

The proposed architecture provided a standard hyper-parameters tuning among the multiple tested data from each station, shown in Table 4.10.

Table 4.10: SSA-window-length-Selection

	Window-Length		
Stations	**S1**	**S2**	**S3**
Benchmark-Models	18	29	16
Proposed-Framework	**9**	**9**	**9**

4.4.5 Discussion and Conclusion

In this framework, a sophisticated wind speed prediction strategy was designed, in order to record the strongest and most reliable wind speeds in Algeria for the best wind parks hosting strategies. For purposes of comparison, two benchmark model classes were chosen.

The application of the GRU, that follows the CNN network, significantly improved the prediction performance and reduced the errors estimation.

In comparison to the CNN-GRU model, the introduction of the ED architecture resulted in a more efficient forecasting and enhanced extraction of features of wind speed among the various collected dataset from the 3 stations choosed. This confirms the effectiveness of the Encoder-Decoder architecture.

The comparison of the three stations revealed that the S1-DAUA station has the most stable annual wind speed distribution, making its variability of wind speed more predictable, having the greatest potential for a new wind park establishment close to this station.

The SSA method used for data denoising demonstrated competitive smoothing results and a significant enhancement in the model performance.

The importance of picking the optimal combination of parameters for the forecasting process was demonstrated by the improved results while using the AGWO in selecting the most efficient parameters combination. This improved the forecasting adaptability and reliability.

The experimental findings demonstrated that the suggested framework surpassed the benchmarks models, with corresponding MAE, MAPE, and RMSE of 0.0081(m/s), 4.2431(%), and 0.0461(m/s), respectively.

Chapter 5

General conclusion and perspectives

”The strongest arguments prove nothing so long as the conclusions are not verified by experience. Experimental science is the queen of sciences and the goal of all speculation.”

-Roger Bacon

”Certainly, there are very few wholly original ideas in the world; there are usually people coming to the same conclusions or carrying out experiments, whether publicly or privately.”

- Shannon Lee

5.1 Conclusion

The main focus of this doctoral dissertation is the creation of an intelligent hybrid framework using different deep learning models combined with sophisticated preprocessing techniques, and apply it in the context of weather forecasting.

This study makes several contributions by combining the fields of DL and weather prediction. While applying the proposed combinations to multiple wind sites, we validated the utility of DL methods for predicting wind speed from various time series data.

We understood the impact of the preprocessing step and the utility of using hybrid combinations, that enabled us to develop more efficient techniques with the best fitting parameters, and reach a clear comprehension of the dynamics in the architecture depth of the wind forecasting process.

The outcomes of the different proposed frameworks showed the superior performance of the suggested architecture designs, reaching 0.0081(m/s), 4.2431(%), and 0.0461(m/s), with corresponding MAE, MAPE, and RMSE, respectively.

The analysis of the various stations confirmed that station S1-DAUA in the Adrar city zone has more stable distribution and higher levels of wind speed throughout the year, which makes the variability of the wind speed more predictable at its level, and shows a great potential for hosting a wind farm near this location.

5.2 Perspectives

(1) Improving the proposed architectures by making new experiments, with various techniques of models hybridization in order to select the more appropriate.

(2) Testing the implemented architectures over different types of data, to validate their flexibility to the changes in its characteristics.

(3) Improving the amount of input series to forecast wind speed with multiple meteorological components, such as pressure, altitude, humidity and wind direction.

(4) Adjusting the designed strategy to long-range forecasts will be significantly helpful in forecasting climate changes.

Bibliography

[1] Javed M.S. et al. Solar and wind power generation systems with pumped hydro storage: Review and future perspectives. *Renewable Energy*, 148: 176–192, 2020.

[2] Wang H. et al. Sparse gaussian process regression for multi-step ahead forecasting of wind gusts combining numerical weather predictions and on-site measurements. *Journal of Wind Engineering and Industrial Aerodynamics*, 220:104873, 2022.

[3] Fathi M. et al. Big data analytics in weather forecasting: A systematic review. *Archives of Computational Methods in Engineering*, pages 1–29, 2021.

[4] Hewage P. et al. Deep learning-based effective fine-grained weather forecasting model. *Pattern Analysis and Applications*, 24(1):343–366, 2021.

[5] Vasquez T. *Weather Forecasting Red Book: Forecasting Techniques for Meteorology*. Weather Graphics Technologies, 2009.

[6] Michalski L. Temperature measurement. *John Wiley Sons*, 2001.

[7] Fente D.N. et al. Weather forecasting using artificial neural network. *second international conference on inventive communication and computational technologies (ICICCT)*, pages 1757–1761, 2018.

[8] Bengtsson L. et al. On the impact of humidity observations in numerical weather prediction. *Tellus A: Dynamic Meteorology and Oceanography*, 57 (5):701–708, 2005.

[9] Sonderby C.K. et al. Metnet: A neural weather model for precipitation forecasting. *arXiv preprint arXiv:2003.12140*, 2020.

[10] Flannery T. The weather makers: The history and future impact of climate change. *Text Publishing*, 2008.

[11] Henson R. Weather on the air: A history of broadcast meteorology. *Springer Science Business Media*, 2013.

[12] Dutta S. et al. Load and renewable energy forecasting for a microgrid using persistence technique. *Energy Procedia*, 143:617–622, 2017.

[13] Su S. et al. Identification of synoptic weather types over taiwan area with multiple classifiers. *Atmospheric Science Letters*, 19(12):e861, 2018.

[14] Bauer P. et al. The quiet revolution of numerical weather prediction. *Nature*, 525(7567):47–55, 2015.

[15] Hamilton J.D. Time series analysis. *Princeton university press*, 2020.

[16] De Gooijer J.G. 25 years of time series forecasting. *International journal of forecasting*, 22(3):443–473, 2006.

[17] Chatfield C. The analysis of time series: an introduction. *Chapman and hall/CRC*, 2003.

[18] Priestley M.B. Non-linear and non-stationary time series analysis. *London: Academic Press*, 1988.

[19] Deb C. et al. A review on time series forecasting techniques for building energy consumption. *Renewable and Sustainable Energy Reviews*, 74: 902–924, 2017.

[20] Jason B. Strategies for multi-step time series forecasting., 2017. URL https://machinelearningmastery.com/multi-step-time-series-forecasting.

[21] Lim B. et al. Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society A*, 379(2194):20200209, 2021.

[22] Taieb S.B. et al. A review and comparison of strategies for multi-step ahead time series forecasting based on the nn5 forecasting competition. *Expert systems with applications*, 39(8):7067–7083, 2012.

[23] Kelleher J.D. Deep learning. *MIT press*, 2019.

[24] Chevillon G. Direct multi-step estimation and forecasting. *Journal of Economic Surveys*, 21(4):746–785, 2007.

[25] Goodfellow I. et al. Deep learning. *MIT press*, 2016.

[26] LeCun Y. et al. Deep learning. *nature*, 2015.

[27] Liu W. et al. A survey of deep neural network architectures and their applications. *Neurocomputing*, 234:11–26, 2017.

[28] Ramachandran P. et al. Searching for activation functions. *arXiv preprint arXiv:1710.05941*, 2017.

[29] Brownlee J. Deep learning for time series forecasting: predict the future with mlps, cnns and lstms in python. *Machine Learning Mastery*, 2018.

[30] Bontempi G. et al. Machine learning strategies for time series forecasting. *European business intelligence summer school,Springer*, pages 62–77, 2012.

[31] Gamboa J.C.B. Deep learning for time-series analysis. *arXiv preprint arXiv:1701.01887*, 2017.

[32] Maiseli B.J. Optimum design of chamfer masks using symmetric mean absolute percentage error. *EURASIP Journal on Image and Video Processing*, (1):1–15, 2019.

[33] Zou Z. et al. Suitability of data preprocessing methods for landslide displacement forecasting. *Stochastic Environmental Research and Risk Assessment*, 34(8):1105–1119, 2020.

[34] Wang Y. et al. A review of wind speed and wind power forecasting with deep neural networks. *Applied Energy*, 304:117766, 2021.

[35] Meenal R. et al. Weather forecasting for renewable energy system: a review. *Archives of Computational Methods in Engineering*, pages 1–17, 2022.

[36] Moriarty P. et al. What is the global potential for renewable energy? *Renewable and Sustainable Energy Reviews*, 16(1):244–252, 2012.

[37] Chabouni N. et al. Electricity load dynamics, temperature and seasonality nexus in algeria. *Energy*, 200:117513, 2020.

[38] Hannah Ritchie and Max Roser. Access to energy, 2020. URL https://ourworldindata.org/energy-access.

[39] Bouznit M. et al. Measures to promote renewable energy for electricity generation in algeria. *Sustainability*, 12(4):1468, 2020.

[40] Nedjari H.D. et al. Optimal windy sites in algeria: Potential and perspectives. *Energy*, 147:1240–1255, 2018.

[41] Fu Y. et al. Multi-step ahead wind power forecasting based on recurrent neural networks. *IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, pages 217–222, 2018.

[42] Devi A.S. et al. Hourly day-ahead wind power forecasting with the eemd-cso-lstm-efg deep learning technique. *Soft Computing*, 24(16): 12391–12411, 2020.

[43] Mehrkanoon S. Deep shared representation learning for weather elements forecasting. *Knowledge-Based Systems*, 179:120–128, 2019.

[44] Hong Y.Y. et al. A hybrid deep learning-based neural network for 24-h ahead wind power forecasting. *Applied Energy*, 250:530–539, 2019.

[45] Hui L. et al. Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition. *Energy Conversion and Management*, 159:54–64, 2018.

[46] Deng Y. et al. A deep learning methodology based on bidirectional gated recurrent unit for wind power prediction. *14th IEEE Conference on Industrial Electronics and Applications (ICIEA) IEEE*, pages 591–595, 2019.

[47] Cebeci Y.E. A recurrent neural network model for weather forecasting. *4th International Conference on Computer Science and Engineering (UBMK)*, pages 591–595, 2019.

[48] Liu H. et al. Smart wind speed deep learning based multi-step forecasting model using singular spectrum analysis, convolutional gated recurrent unit network and support vector regression. *Renewable energy*, 143:842–854, 2019.

[49] Mishra S. et al. Comparison of deep learning models for multivariate prediction of time series wind power generation and temperature. *Energy Reports*, 6:273–286, 2020.

[50] Zhou Q. et al. Hybrid forecasting system based on an optimal model selection strategy for different wind speed forecasting problems. *Applied Energy*, 250:1559–1580, 2019.

[51] Ambach D. et al. Space-time short-to medium-term wind speed forecasting. *Statistical Methods and Applications*, 25(1):5–20, 2016.

[52] Liu Z. et al. A combined forecasting model for time series: Application to short-term wind speed forecasting. *Appl Energy*, 259:114137, 2020.

[53] Jiang Z. et al. Ultra-short-term wind speed forecasting based on emd-var model and spatial correlation. *Energy Conversion and Management*, 250: 114919, 2021.

[54] Wang Y. et al. Approaches to wind power curve modeling: A review and discussion. *Renew Sustain Energy Rev*, 116(109422), 2019.

[55] Liu H. et al. Data processing strategies in wind energy forecasting models and applications: A comprehensive review. *Appl Energy*, 249:392–408, 2019.

[56] Feature extraction, 2016. URL https://deepai.org/machine-learning-glossary-and-terms/feature-extraction/.

[57] Han J. Data mining: concepts and techniques. *Micheline KAMBER a Jian PEI*, pages 6–7, 2012.

[58] Jiawei H. *Data Preprocessing in Data Mining Concepts and Techniques Third Edition*. Elsevier,Chapter 3,pp. 83-124, 2012.

[59] Li Y. et al. Bayesian robust multi-extreme learning machine. *Knowl-Based Syst*, 210:106468, 2020.

[60] Fu Q. et al. Multi-stations' weather prediction based on hybrid model using 1d cnn and bi-lstm. *Chinese control conference IEEE*, pages 3771–3775, 2019.

[61] CFI Team. Standardize function, 2022. URL https://corporatefinanceinstitute.com/resources/excel/z-score-standardize-function/.

[62] Codecademy Team. Normalization, 2022. URL https://www.codecademy.com/article/normalization/.

[63] Guo Y. et al. Multi-step forecasting for wind speedusing a modified emd-based artificial neural network model. *Renew Energy*, 37(1):241–9, 2012.

[64] Ellis C.A. et al. Is smarter better? a comparison of adaptive, and simple moving average trading strategies. *Research in International Business and Finance*, 19(3):399–411, 2005.

[65] Cadenas E. et al. Analysis and forecasting of wind velocity in chetumal, quintana roo, using the single exponential smoothing method. *Renewable Energy*, 35(5):925–930, 2010.

[66] Cambron P. et al. Power curve monitoring using weighted moving average control charts. *Renewable Energy*, 94:126–135, 2016.

[67] Oh E. et al. Energy-storage system sizing and operation strategies based on discrete fourier transform for reliable wind-power generation. *Renewable Energy*, 116:786–794, 2018.

[68] Jaseena K.U. et al. Decomposition-based hybrid wind speed forecasting model using deep bidirectional lstm networks. *Energy Conversion and Management*, 234:113944, 2021.

[69] Jaseena L. et al. Lecture notes on wavelet transforms. *Boston: Birkhauser*, 2017.

[70] Brusco S. et al. Thunderstorm-induced mean wind velocities and accelerations through the continuous wavelet transform. *Journal of Wind Engineering and Industrial Aerodynamics*, 221:104886, 2022.

[71] Jnr E.O.N. et al. A hybrid chaotic-based discrete wavelet transform and aquila optimisation tuned-artificial neural network approach for wind speed prediction. *Results in Engineering*, 14:100399, 2022.

[72] Huang N. et al. Short-term wind speed forecast with low loss of information based on feature generation of osvd. *IEEE Access*, 7:81027–81046, 2019.

[73] Hu H. et al. Wind speed forecasting based on variational mode decomposition and improved echo state network. *Renewable Energy*, 164: 729–751, 2021.

[74] Liu H. et al. Smart multi- step deep learning model for wind speed forecast- ing based on variational mode decomposition, sin- gular spectrum analysis, lstm network and elm. *Energy Conversion and Management*, 159: 54–64, 2018.

[75] Mi X. et al. Wind speed prediction based on singular spectrum analysis and neural network structural learning. *Energy Conversion and Management*, 216:112956, 2020.

[76] Durbin J. Efficient estimation of parameters in moving-average models. *Appl Energy*, 46(3/4):306–316, 1959.

[77] Sneddon I.N. Fourier transforms. *Courier Corporation*, 1995.

[78] Brigham E.O. *The fast Fourier transform and its applications*. Prentice-Hall, Inc., 1988.

[79] Rao K.R. et al. Fast fourier transform: algorithms and applications. *Dordrecht: Springer*, 32, 2010.

[80] Dragomiretskiy K. et al. Variational mode decomposition. *IEEE transactions on signal processing*, 62(3):531–544, 2013.

[81] Liao X. et al. Short-term wind speed multistep combined forecasting model based on two- stage decomposition and lstm. *Wind Energy*, 24(9): 991–1012, 2021.

[82] Lian J. et al. Adaptive variational mode decomposition method for signal processing based on mode characteristic. *Mechanical Systems and Signal Processing*, 107:53–77, 2018.

[83] Wang C. et al. Wind power forecasting based on singular spectrum analysis and a new hybrid laguerre neural network. *Applied Energy*, 259: 114139, 2020.

[84] Golyandina N. et al. Singular spectrum analysis with r. *Springer Berlin Heidelberg*, pages 273–286, 2018.

[85] Li W. et al. A survey of learning-based intelligent optimization algorithms. *Archives of Computational Methods in Engineering*, 28(5):3781–3799, 2021.

[86] Liashchynskyi P. et al. Grid search, random search, genetic algorithm: a big comparison for nas. *arXiv preprint arXiv:1912.06059*, 2019.

[87] Sen S.Y. et al. Convolutional neural network hyperparameter tuning with adam optimizer for ecg classification. *Innovations in Intelligent Systems and Applications Conference (ASYU),IEEE*, pages 1–6, 2020.

[88] Mirjalili S. et al. Grey wolf optimizer. *Advances in Engineering Software*, 69:46–61, 2014.

[89] Nadimi-Shahraki M. H. et al. An improved grey wolf optimizer for solving engineering problems. *Expert Systems with Applications*, 166, 2021.

[90] Gu J. et al. Recent advances in convolutional neural networks. *Pattern recognition*, 77:354–377, 2018.

[91] Vedaldi A. et al. Matconvnet: Convolutional neural networks for matlab. *Proceedings of the 23rd ACM international conference on Multimedia*, pages 689–692, 2015.

[92] Nair V. et al. Rectified linear units improve restricted boltzmann machines. *Icml*, 2010.

[93] Eldar Y.C. et al. Robust mean-squared error estimation in the presence of model uncertainties. *IEEE Transactions on Signal Processing*, 53(1): 168–181, 2004.

[94] Liu Y. et al. Wind power short-term prediction based on lstm and discrete wavelet transform. *Applied Sciences*, 9(6):1108, 2019.

[95] Altan A. et al. A new hybrid model for wind speed forecasting combining long short-term memory neural network, decomposition methods and grey wolf optimizer. *Applied Soft Computing*, 100:106996, 2021.

[96] Chen G. et al. Short-term wind speed forecasting based on long short-term memory and improved bp neural network. *International Journal of Electrical Power Energy Systems*, 134:107365, 2022.

[97] Chung J. et al. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014.

[98] Peng C. et al. Multi-step-ahead host load prediction with gru based encoder-decoder in cloud computing. *10th International Conference on Knowledge and Smart Technology (KST),IEEE*, pages 186–191, 2018.

[99] Shewalkar A. et al. Performance evaluation of deep neural networks applied to speech recognition: Rnn, lstm and gru. *Journal of Artificial Intelligence and Soft Computing Research*, 9(4):235–245, 2019.

[100] Jaseena K.U et al. Deterministic weather forecasting models based on intelligent predictors: A survey. *Journal of King Saud University-Computer and Information Sciences*, 2020.

[101] Grover A. et al. A deep hybrid model for weather forecasting. *21th ACM SIGKDD international conference on knowledge discovery and data mining*, pages 379–386, 2015.

[102] Zouaidia K. et al. Hourly wind speed forecasting using fft-encoder-decoder-lstm in south west of algeria (adrar). *International Journal of Informatics and Applied Mathematics*, 4(1):72–83, 2021.

[103] Ltd.Weather. Raspisaniye pogodi ltd.weather for 243 countries of the world, 2020. URL https://rp5.ru/Weather_in_the_world.

[104] Zouaidia K. et al. Wind speed forecasting based on discrete wavelet transform, moving average method and gated recurrent unit. *International Conference in Artificial Intelligence in Renewable Energetic Systems, Springer*, pages 71–78, 2020.

[105] Zouaidia K. et al. Multi-step wind speed forecasting based on hybrid deep learning model and trailing moving average denoising technique. *International Conference on Recent Advances in Mathematics and Informatics (ICRAMI), IEEE*, pages 1–5, 2021.

[106] Global wind atlas, 2022. URL https://globalwindatlas.info/en/area/Algeria/.

[107] Zouaidia K. et al. Hybrid intelligent framework for one-day ahead wind speed forecasting. *Neural Computing and Applications*, 33(23): 16591–16608, 2021.