Ministry of Higher Education and Scientific Research

وزارة التعليم العالي والبحث العلمي

Badji Mokhtar Annaba University
Université Badji Mokhtar – Annaba
Faculté de Technologie

Département Electronique

جامعة باجي مختار– عنابـــة

كـليـة التكنـولـوجيـا

قســم الإلكترونيك

Thesis

Submitted to obtain the degree of

Doctorate Third Cycle

Field : Electronics

Speciality : Telecommunications

By :

CHELABI Hiba

Title :

Apprentissage des Réseaux de Neurones Profonds sur des Données Multimodales

Thesis defended on Dec 2nd 2025 in front of the committee composed of:

N°	Full name	Rank	Establishment	Quality
01	Doghmane Noureddine	Prof.	Badji Mokhtar Annaba University	President
02	**Khadir Mohamed Tarek**	Prof.	Badji Mokhtar Annaba University	Supervisor
03	Chikhaoui Belkacem	Prof.	Teluq University	Co-supervisor
04	Farah Nadir	Prof.	Badji Mokhtar Annaba University	Examiner
05	Azizi Nabiha	Prof.	Badji Mokhtar Annaba University	Examiner
06	Farou Brahim	Prof.	8 mai 1954 Guelma University	Examiner

First and foremost, there is no doubt that this is good luck from God, who guided me along this path to be where I am today and accomplish this research work.

Furthermore, this achievement would not have been possible without my determination towards this research work as well as the contribution of several people, including my parents, to whom I express my sincere gratitude.

Acknowledgment

Firstly and more particularly, I would like to express my deep gratitude to my Ph.D. thesis director, Pr. **Mohamed Tarek Khadir**, Professor at the University of Badji Mokhtar Annaba, who gave me intellectual freedom in my work, guided me throughout each stage of this research journey, and always inspired my thoughts with his enthusiastic encouragement, which consequently made this work flow with ease.

I also would like to thank my thesis co-director, Pr. **Belkacem Chikhaoui**, Professor at TÉLUQ University, who considerably supported me from behind the scenes. His availability and patience have been instrumental in the success of this work.

Not to forget the members of the LabGED laboratory and professors at the University of Badji Mokhtar Annaba, who provided an atmosphere of trust and support.

Last but not least, I would like to express my deep sense of gratitude to the support system that has always been by my side: **My family**.

I would like to extend my thanks to the members of the thesis committee for the time and effort given to examine this dissertation, Pr. **Nabiha Azizi**, Pr. **Nadir Farah**, Pr. **Noureddine Doghmane** and Pr. **Brahim Farou**. Their invaluable feedback and insights will help to improve and refine my research insights and methodology.

«تعلم الشبكات العصبية العميقة على البيانات متعددة الوسائط»

الملخص

إن إدراكنا للعالم يكون في أغلب الأحيان متعدد الوسائط. بعبارة أخرى، لا يقتصر الذكاء الطبيعي على نمط واحد فقط. إن دماغنا قادر على معالجة أنواع متعددة من المعلومات ذات الصلة بمهمة محددة في وقت واحد، وينطبق نفس الشيء على الذكاء الاصطناعي. بشكل عام، يحمل دمج الوسائط المتعددة وعدًا كبيرًا بتطوير قدرات الذكاء الاصطناعي في معالجة المعلومات متعددة الوسائط.

تناقش هذه الأطروحة بشكل خاص آليات الوسائط المتعددة في سياق الذكاء الاصطناعي، مثل دمج المعلومات متعددة الوسائط من وسائط استشعار مختلفة وتشير إلى أهمية الإدراك متعدد الوسائط.

يتم تدريب وتقييم نماذج التعلم العميق المختلفة باستخدام بيانات مقدمة من شركة الكهرباء والغاز الوطنية SONELGAZ. ويُقيّم أداء المساهمات المختلفة بمقارنة النتائج التجريبية لمختلف البنى العميقة.

كلمات مفتاحية: الوسائط التعددية، الاندماج متعدد الوسائط، التعلم العميق، الشبكات العصبية الاصطناعية، التنبؤ باستهلاك الكهرباء على المدى القصير.

«Apprentissage des réseaux de neurones profonds sur des données multimodales»

Résumé

Notre perception du monde est la plupart du temps multimodale. En d'autres termes, l'intelligence naturelle ne se limite pas à une seule modalité. Notre cerveau est capable de traiter simultanément plusieurs types d'informations pertinentes pour une tâche spécifique, et il en va de même pour l'intelligence artificielle. Dans l'ensemble, l'intégration de la multimodalité est très prometteuse pour faire progresser les capacités de l'intelligence artificielle dans le traitement des informations multimodales.

Cette thèse traite en particulier des mécanismes de la multimodalité dans le contexte de l'intelligence artificielle, c'est-à-dire de la fusion d'informations multimodales provenant de différentes modalités de capteurs, et souligne l'importance de la perception multimodale.

Différents modèles d'apprentissage profond sont formés et évalués à partir des données fournies par la compagnie nationale d'électricité et de gaz SONELGAZ. Les performances des différentes contributions sont évaluées en comparant les résultats empiriques des différentes architectures profondes.

Mots clés : Multi-modalité, fusion multi-modale, apprentissage profond, réseaux de neurones artificiels, prévision de la charge électrique à court terme.

«Training deep neural networks on multimodal data»

Abstract

Our perception of the world is most of the time multi-modal. In other words, natural intelligence is not limited to just a single modality. Our brain is able to process multiple types of information relevant to a specific task simultaneously, and the same applies to artificial intelligence. Overall, the integration of multi-modality holds great promise for advancing the capabilities of artificial intelligence in processing multi-modal information.

This thesis discusses particularly the mechanisms of multi-modality in the context of Artificial Intelligence i.e. such as merging multi-modal information from different sensor modalities and points to the importance of multi-modal perception.

Different Deep Learning models are trained and evaluated using data provided by the national electricity and gas company SONELGAZ. The performance of the various contributions is evaluated by comparing the empirical results of the different deep architectures.

Key words: Multi-modality, Multi-modal Fusion, Deep Learning, Artificial Neural Networks, Short-term Load Forecasting.

Contents

List of Figures

List of Tables

List of Acronyms

1D One Dimensional.

2D Two Dimensional.

3D Three Dimensional.

AE Auto-encoders.

AF Activation Function.

AI Artificial Intelligence.

ANN Artificial Neural Network.

Bi-GRU Bidirectional Gated Reccurent Unit.

Bi-LSTM Bidirectional Long Short Term Memory.

CNN Convolutional Neural Networks.

CNN-1D One Dimensional Convolutional Neural Networks.

CNN-2D Two Dimensional Convolutional Neural Networks.

DAE Denoising Auto-encoders.

DANN Deep Artificial Neural Network.

DBN Deep Belief Network.

DL Deep Learning.

FC Fully Connected.

GRU Gated Reccurent Unit.

LSTM Long Short Term Memory.

LTLF Long-term load Forecasting.

MAE Mean Absolute Error.

MAPE Mean Absolute Percentage Error.

ML Machine Learning.

MLP Multi-layer Perceptron.

MSE Mean Squared Error.

MTLF Medium-term load Forecasting.

RL Reinforcement learning.

RMSE Rooted Mean Squared Error.

RNN Reccurent Neural Network.

SAE Stacked Auto-encoders.

SDAE Stacked Denoising Auto-encoders.

SL Supervised Learning.

SONELGAZ SONELGAZ.

SOTA State of the Art.

SSL Semi-supervised Learning.

STLF Short-term Load Forecasting.

STPF Short-term Price Forecasting.

SVM Support Vector Machine.

UL Unsupervised Learning.

VMD Variational Mode Decomposition.

General Introduction

AI is a multidisciplinary field that consists of the implementation of techniques allowing machines to simulate a form of human intelligence. Human intelligence, on the other hand, is a mental quality that consists of integrating the human senses in synchrony to understand the world around us, learn from experiences and adapt to new situations.

There are five basic human senses that work in coordination: sight, touch, hearing, smell and taste. For instance, human speech is likely supplemented by a range of modes that may arise in combination and/or independently adding clarity to the information being conveyed, such as gestures, tone and facial expressions.

The term that replaces the human senses in the context of AI is "modality". It is classified as a single independent channel of sensory input. A system is therefore designed as uni-modal if it includes one modality and multi-modal otherwise.

Multi-modal systems aim to imitate the cognitive abilities of human beings by combining different modalities of information simultaneously to perform different tasks.

In the last decade, much progress that has been made in the AI research field was attributable to ANN algorithms, backed with DL libraries that are designed to make the most accurate predictions possible.

The wide spectrum of architectures that represent deep learning has made an extensive development in the field of artificial intelligence, allowing us to build solutions for a range of problems in different areas.

These solutions such as; classification, segmentation, dimensionality reduction etc. are mostly based on single modalities.

That is the reason why we must innovate further and put our focus on how to exploit the strength of the present methods when combined together to obtain the best results while processing information from multiple modal-

ities, and exploring the potential of multimodal deep learning. Starting from the hypothesis that regardless of whether the response to each modality on its own is weak or even strong, then the opportunity for accuracy enhancement is very large when it comes to integrating multiple modalities to collaborate in performing the same task. Whence comes the term multi-modality in AI, which can be defined as the depiction or communication of one or multiple ideas using more than a single expressive mode, either in synchrony or separately [1].

Different deep learning applications were subject to tremendous studies in the literature based on multi-modal data, and few of these applications are related to the electrical grid [2, 3].

Grid technology is in active development around the world as it tops the list of life-changing innovations. It continues to expand dramatically even as people look for other energy sources to displace it. The use of electricity for operating modern technologies on a daily basis, starting from domestic use to industrial activities, is clearly seen as the need of reliable and continuous energy. It is very much understandable as it is a major economic growth and development factor. Hence, smart grid technology is under continuous development to improve efficiency in how power is monitored and metered for each customer.

In fact, a high per capita electricity consumption in developed countries is also correlated with high numbers on the human development index. Therefore, numerous studies demonstrate the connection between economic growth and electricity consumption levels [4, 5].

Problem Statement

Load forecasting ensures the availability of enough power to meet consumption needs while avoiding waste and inefficiency. This research field has never failed to attract researchers' interest, and to keep abreast of developments, we extend it to a broader research field called multi-modal AI. The intersection of these two fields aims to unlock new frontiers in AI and innovate further with scientific research.

The core of this thesis seeks to answer the questions of: How to combine the different modalities within a DL model? And what are the benefits of combining multi-modal data that would not have been present if the processing were only uni-modal?

In the present thesis, we tried to tackle the STLF task based on multi-modal data as it holds great promise for advancing the capabilities of ANN models in processing multi-modal information.

Thesis Outline

This Thesis is structured as follows:

We shed light on the intersection between AI, ML, DL, ANN in chapter 1. We also covered the fundamental theories behind several DANN, which are employed in the literature for uni-modal processing and which we will subsequently apply to the processing of multi-modal data. In chapter 2, the context of multi-modality in AI is elucidated. Furthermore, on the one hand, the different types of modalities as well as the taxonomy of multi-modal AI have been reviewed. chapter 3 on the other hand, grouped the existing methods and applications in multi-modal AI with relevance to multiple research areas in order to highlight recent advances and trends.

In chapter 4, the focus was put on reviewing the SOTA methods applied to load forecasting based on uni-modal data. In chapter 5, the Algerian electric load forecasting task has been tackled based on uni-modal data. After reviewing the load forecasting works based on uni-modal data as well as tackling the load forecasting based on uni-modal data, DL methods have been applied in chapter 6 to analyse their impact of STLF. chapter 7 is dedicated to exploit the strength of one of the multi-modal research directions "multi-modal fusion" when combining different DL methods as well as different sources of information.

Chapter 1

Artificial Intelligence Towards Advanced Machine Learning Techniques and Algorithms

1.1 Introduction

There is a spacious field that encompasses ML, DL and ANN, with the aim to replicate the human brain called AI. The birth of the term AI was at the Dartmouth Summer Research Project on AI in 1956, where John McCarthy stated that the conference was to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it [6] which was widely considered to be the founding event of AI as a field.

In the most basic sense, the term AI for short makes reference to the simulation of human-level intelligence that relies on the creation and application of algorithms executed in a dynamic computing environment to solve a wide variety of tasks smartly. It involves ML and DL techniques to execute these tasks so much so that these terms are often used interchangeably; however, we should not confuse between them.

To narrow it down in a single sentence: DL is a specialized subset of ML, which in turn is a subset of AI.

AI is a broad field that covers tremendous research areas like Natural Language Understanding, Computer Vision, Recommendation systems, etc. This is to mention a few and in broad terms, but specific advances of AI

nowadays in real life are for instance; Self-driving cars (Tesla), smart assistants (Siri), chatbots (chatGPT), etc.

With our continuous quest for faster, smarter and more accurate ways of getting things done, AI is quickened at a phenomenal rate, which in turn lays in DL techniques that brought AI close to its purpose.

In the present chapter, we want to clarify how AI, ML, ANN and DL intersect. We will present as well the core idea behind DANN architectures that are used in the literature to perform the uni-modal processing.

1.2 Artificial Intelligence

When they think of AI, many people think of the robots that act like human beings that they have seen in the movies, but that is not the current reality of the field. AI, which is everywhere in our day to day lives, is a technology focused on performing one specific task or a narrow range of tasks. It is more properly called Artificial Narrow Intelligence, or Narrow AI. its goal is to develop systems that can perform tasks that typically require human intelligence.

Narrow AI can only perform the task it was designed to do. It doesn't have the flexibility to work in a different context than the one it was designed for. A few examples of Narrow AI includes spam filters, autonomous cars, recommendation systems, object detection, etc.

Basically, an AI system is a collection of mathematical algorithms that make computers make sense of big amounts of data by implementing models that understand complex relationships among this data, automate processes and tasks and make actionable decisions.

1.3 Machine Learning

ML algorithms often require handcrafted feature engineering, where human experts design the input features that the model uses for learning and prediction.

ML is at the core of the definition of AI, while the latter aims to give a machine the ability to reason and behave like a human, ML is only one way to help move towards this vision with less human intervention. It is a form of AI that focuses on creating systems that learn and improve their

performance from experience using a training set of examples, which means, based on the data they process, the algorithm learns from its error and refines itself. ML algorithms often require handcrafted feature engineering, where human experts design the input features that the model uses for learning and prediction. Similarly to the human brain, it relies on inputs to gain knowledge, understand and detect trends in a knowledge history, for example, it can use the temperature factor to derive a prediction model for the future, here is the weather or the electricity consumption.

ML can be categorized into several types based on the learning approach and the nature of the tasks; SL, UL, SSL and RL.

1.3.1 Supervised Learning

SL also known as Supervised ML, is a subcategory of ML where the algorithm learns a mapping or relationship between input features (X) and output labels (Y) Eq. (1.1), so that it can generalize and hence make predictions on unseen data. The term "supervised" refers to the process of guiding the algorithm by providing it with a labeled training set.

$$Y = f(X) \tag{1.1}$$

In SL, the training data-set must include inputs and corresponding outputs to allow the model to learn over time. A loss function is used to quantify the difference between the predicted values and the actual outputs. The goal is to minimize this loss during training.

SL can be categorised into two types of problems Classification and Regression:

Classification

Classification is a ML technique used to categorize a given data sample into one or more predefined classes or categories. The predicted classes are discrete labels, but a classification algorithm may predict a continuous value that presents a probability for a class label. Some common applications of classification include image classification, spam filtering, sentiment analysis etc.

Common classification models are the following: Logistic Regression, Naive Bayes, k-Nearest Neighbor, Decision Trees, Support Vector Machine and Neural Networks.

Regression

Regression, on the other hand, is the process of modeling the relationship between input and output variables referred to as dependent and independent variables, respectively, where the output is in the form of a continuous numerical value or quantity. Its goal is to find the mathematical equation that best describes this relationship so that predictions based on new inputs could be made.

Common regression models are the following: Linear Regression, Polynomial Regression, Decision Trees, Neural Networks.

1.3.2 Unsupervised Learning

Unlike SL, UL models are used when the training data is neither classified nor labeled. From that data, the task is to discover patterns that help to group the information according to similarities and differences even though no categories are provided.

Clustering

Clustering algorithms group similar instances of the data together in one cluster based on certain features, so that the intra-cluster similarities are maximized while the inter-cluster similarities are minimized. Some real world applications of clustering include categorizing books in a library, anomaly detection, customer segmentation for marketing purposes.

Common clustering models are the following: K-means, BIRCH, OPTICS.

Dimensionality Reduction

Dimensionality reduction techniques aim to reduce the number of input features while preserving the essential information in the data. This can be useful for visualizing high-dimensional data or speeding up subsequent ML algorithms, it is also considered as a data preprocessing step.

Common dimensionality reduction models are the following: Principal component analysis, autoencoders.

1.3.3 Semi-supervised Learning

SSL is a ML paradigm that lies between SL and UL. In SSL, the algorithm is trained on a data-set that contains a combination of labeled and unlabeled examples. This approach is particularly useful, as it can be time consuming and costly to rely on domain expertise to label data appropriately for SL. The goal is to take advantage of the large amounts of unlabeled data available to improve the accuracy of the model while still incorporating the labeled data to provide proper supervision during training.

Common semi-supervised learning methods are the following: semi-supervised SVM, Manifold regularization, Manifold approximation, Generative Adversarial Networks.

1.3.4 Reinforcement Learning

RL is a type of ML paradigm where an agent learns to make decisions by interacting with an environment. The agent takes actions, receives feedback in the form of rewards or penalties, and adjusts its strategy to maximize cumulative reward over time. In other words, the agent learns to take actions in an environment to achieve a desired outcome. RL is particularly well-suited for problems where explicit training data is scarce, and the agent must learn from its own experiences.

Common reinforcement learning methods are the following: Q-learning, DDPG, PPO.

1.4 Deep Learning

DL is a more evolved branch of ML, it is paired more specifically with the concept of ANN as they make up the backbone of any DL algorithm. The latter uses a complex structure modeled on the human brain which enables the processing of unstructured data such as images, documents and text. As previously stated, DL method is based on ANN, and more specifically, DANN,

A DANN consists of a hierarchy in the form of algorithms that processes data through several layers of weighted nodes to make a decision. The outcome of each layer are fed as input to the next layer. Unlike ML, there is no direct human involvement in the model, all the feature extraction is done by the algorithm.

A DL model does not simply follow instructions provided by a programmer, it makes predictions and recommendations based on patterns and representations learned from data automatically by the algorithm, thereby saving a great deal of effort and time coping with far more complex phenomena. By design, a DL model improves itself with practice, it may evolve beyond the understanding of the person who created it, and it can process massive amounts of data efficiently, but at a high computational cost.

The key concept of to DANN include training. Training a DL model involves presenting it with labeled input data and adjusting the model's parameters (weights and biases) to minimize the difference between the predicted output and the actual output. This is typically done using back-propagation and the optimization algorithm.

1.4.1 Back-propagation Algorithm

When learning about ANN, we come across two essential terms describing the movement of information; Feed-forward and back-propagation. Feed-forward or forward propagation is the way to move from the input layer to the output layer in the ANN to calculate the activation of each neuron, while the reverse process of moving from the output to the input layer is called the backward propagation or back-propagation.

Back-propagation is an algorithm used for computing the gradient of the cost function for each weight in the ANN, which in turn will be used by an optimization algorithm to adjust the model's weights and biases in a direction that reduces the cost.

Actually, back-propagation refers only to the method for calculating the gradient, while the process of using back-propagation along with an optimization algorithm for instance stochastic gradient descent is one of the key algorithms in DL and it is what we call training an ANN with the aim to minimize the loss of a predictive model with regard to a training data-set [7].

The back-propagation algorithm can be summarized as follows:

- Initialize the weights and biases of the network randomly.

- Feed the input through the network and compute the predicted output.
- Compute the error between the predicted output and the true output.
- Compute the gradients of the cost function with respect to the weights and biases in the network.
- Update the parameters of the network by subtracting a fraction of the gradients from the previous step.
- Repeat steps 2 through 5 until the cost converges to a minimum value, or a maximum number of iterations is reached.

1.4.2 Activation Functions

An AF also referred to as a transfer function, is a simple mathematical operation that transforms the summed weighted input from the node into an output value to be fed to the next hidden layer or to the output layer, in other terms it controls the flow of information in the ANN by deciding whether a neuron should be activated or not, this means that it helps in distinguishing the important information from the irrelevant ones.

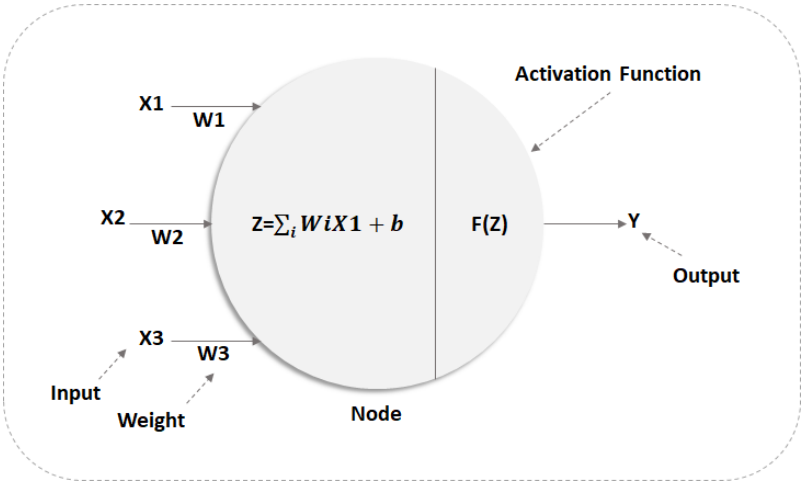


Figure 1.1: Process of Node Activation Function.

An ANN without an activation function is essentially just a linear regression model, that is why activation functions are as important as any part of

the ANN because they add the non-linearity that the network needs in order to be capable to perform more complex tasks. There are several commonly used activation functions such as; linear, Softmax, TanH, RelU and Sigmoid. Each of which have a specific purpose. For a binary classification for example, Sigmoid and Softmax functions are preferred and for a multi-class classification, generally Softmax is used.

The three major types of activation functions are as listed below:

Binary Step Function

Binary step function is a function that produces a binary output, depending on a threshold value, that decides whether a neuron should be activated or not. It can be used while creating a binary classifier, its mathematical presentation is as follows:

$$f(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ \\ 0, & \text{if } x < 0 \end{cases} \tag{1.2}$$

Linear Function

The linear activation function, also known as Identity Function, as its name suggests is where the activation is identical to the input, it can presented mathematically as:

$$f(x) = x \tag{1.3}$$

Non-Linear Function

The significance of the activation function lies in making a given model learn and execute difficult tasks, here comes the importance of non-linear functions, the most common and most used activation functions are:

Sigmoid

This function takes any real value as input and outputs values in the range of 0 to 1. The larger the input (more positive), the closer the output

value will be to 1.0, whereas the smaller the input (more negative), the closer the output will be to 0.0, as shown in equation Eq. (1.4)

$$f(x) = \frac{1}{1 + e^{-x}} \tag{1.4}$$

Hyperbolic Tangent

Hyperbolic Tangent, or (Tanh), for short is a function that is very similar to the sigmoid/logistic activation function, and even has the same S-shape with the difference in output values that range between -1 and 1. In Tanh, the larger the input (more positive), the closer the output value will be to 1.0, whereas the smaller the input (more negative), the closer the output will be to -1.0, see equation Eq. (1.5)

$$f(x) = \frac{(\mathrm{e}^{\mathrm{x}-\mathrm{e}^{-\mathrm{x}}})}{(\mathrm{e}^{\mathrm{x}+\mathrm{e}^{-\mathrm{x}}})} \tag{1.5}$$

Rectified Linear Unit

ReLU stands for Rectified Linear Unit, it is an activation function that looks and acts like a linear function, but is in fact nonlinear allowing complex relationships in the data to be learned. The rectified linear activation function uses a simple calculation that returns the value provided as input directly, or 0 if the input is 0 or less, see equation Eq. (1.6)

$$f(x) = max(0, x) \tag{1.6}$$

Softmax

The Softmax activation function transforms the raw outputs of the neural network into a vector of probabilities, whose total sums up to 1. Consider a multi-class classification problem with N classes, the softmax activation returns an output vector that is N entries long, with the entry at index i corresponding to the probability of a particular input belonging to class i.

$$Softmax(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \quad for\ i = 1, 2, \ldots, K \tag{1.7}$$

Advances in model architectures, training techniques, and hardware have contributed to the success and popularity of DANN in various domains, including computer vision (e.g., image classification and object detection), natural language processing (e.g., language translation and sentiment analysis), speech recognition, and more.

1.4.3 Multi-layer Perceptron

The perceptron was invented in 1957 at the Cornell Aeronautical Laboratory [8]. The first ANN was not able to solve nonlinear problems, this limitation was removed by back-propagation of the error gradient and the optimization algorithms in MLP models [9].

MLP may be considered the most basic ANN. It is made up of an input layer, one or more hidden layers, and an output layer. The information moves forward, from the input layer through the hidden layers, then to the output layer, where the latter returns a decision or prediction about the input. Then the optimization algorithm with the back-propagation performs a backward pass to adjust the model's parameters according to the loss value to help produce a correct input-output mapping.[10]. MLP is commonly used for tasks such as classification and regression.

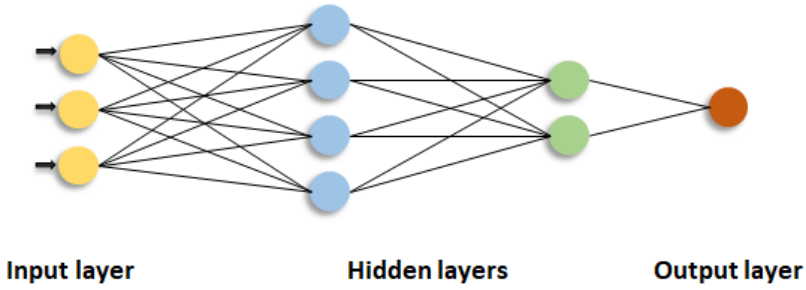


Figure 1.2: Overview of the Multi-layer Perceptron Architecture.

1.4.4 Auto-encoders

Vanilla Auto-encoders

Vanilla Auto-encoders proposed by Hinton in 1980s [9, 11], are a type of DL Algorithms that are used for UL, especially for feature extraction and dimensionality reduction.

An Auto-encoder consists of one hidden layer where the number of neurons in the hidden layer is less than the number of neurons in the input and output layers, therefore we can think of the hidden layer as a bottleneck layer. The process is to transform a high-dimensional input into a low-dimensional one called "latent code", and then a reverse process called "decoder" reconstructs the original data from the latent code [12], this process reduces the amount of data, while still accurately describe the original data-set.

An Auto-encoder takes an input $x \in [0, 1]$ and first encode it to a hidden representation $y \in [0, 1]$ through a non-linearity s such as Sigmoid Eq. (5.2).

$$y = s(wx + b) \tag{1.8}$$

The result y is then decoded back into a reconstruction z as close as possible to the original input x Eq. (1.9), [12].

$$z = s(w'y + b') \tag{1.9}$$

Denoising Auto-encoders

The low-dimensional representation of the data captures as much information as possible from the input, however we need to ensure that the DANN does something more interesting than merely learning the identity of the input, here comes the DAE a variant of Auto-encoders where an amount of noise is added to the input, in order to force the hidden layer to extract more robust features from the input and lead to improvements in network generalization. This variant of Auto-encoders is trained to reconstruct a clean data from its corrupted version by randomly turning some of the input values to zero [13]. In effect, injecting noise in the input of an ANN can also be seen as a form of data augmentation, because it expands the size of the training data-set. Each time a training sample is exposed to the model, a random noise is added to the input variables making them different every time it is exposed to the model [14].

Stacked Denoising Auto-encoders

Working in much the same way as Deep Belief Networks [15], Auto-encoders and DAE can be stacked to form SAE and SDAE respectively, using initially a local unsupervised criterion to pre-train each layer in turn, then the network goes through a second stage of supervised training called fine-tuning, where the network is trained in exactly the same way as a typical MLP, by considering only the encoding parts of each Auto-encoder or DAE and adding an output layer on the top of these stacked layers [13]. The parameters of the whole system are adjusted to minimize the error in predicting the supervised target by performing the back-propagation with the optimization algorithm.

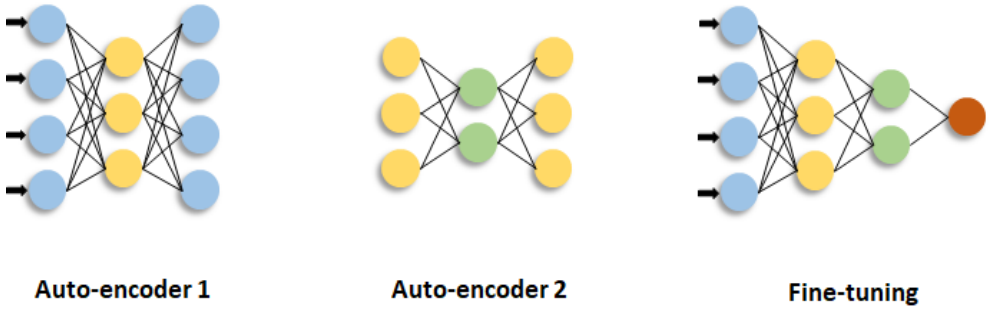


Figure 1.3: Overview of the Stacked Denoising AEs Architecture.

1.4.5 Convolutional Neural Networks

CNN are a type of DANN that were introduced in 1989 specifically for zip code recognition [16]. CNN have become dominant in computer vision, as they achieved SOTA performance on a wide range of tasks, including object detection, image classification, and segmentation.

CNN are known to work on two-dimensional data, however, there are two other types; one-dimensional CNN and three-dimensional CNN, where every single type of CNN accepts a specific structure of input data.

One Dimensional Convolutional Neural Networks

In CNN-1D the kernel is one dimensional, it slides along the input to capture properties and calculate the output which is also in the shape of 1D array, see Eq. (1.10) [17].

$$y(n) = \begin{cases} \sum_{i=0}^{k} x(n+i)h(i), & \text{if n=0.} \\ \sum_{i=0}^{k} x(n+i+(s-1))h(i), & \text{otherwise.} \end{cases} \tag{1.10}$$

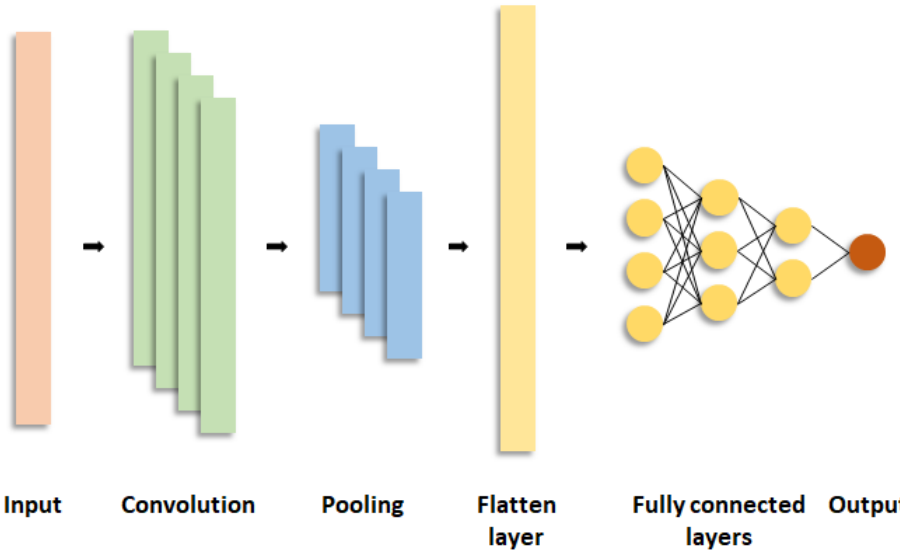


Figure 1.4: Overview of the One-Dimensional Convolutional Neural Network Architecture.

Two-Dimensional Convolutional Neural Network

CNN-2D are designed to operate on two dimensional data such as 2D images, where the kernel is in the form of a small matrix of weights that moves in 2 directions (X,Y) to extract the spatial features from the data such as edges and textures. Eq. (1.4.5) below presents the mathematical operation of the 2D convolution.

$$(f * g)(x, y) = \sum_{s=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} f(s, t)g(x - s, y - t) \tag{1.11}$$

Where:

f is the input image, g is the convolution kernel or filter and (x, y) is the current pixel location in the output feature map. s and t are the pixel indices of the input image.

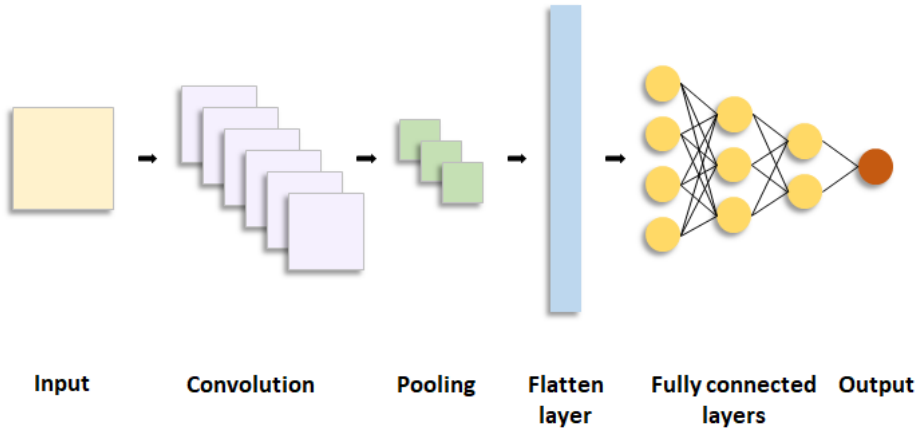


Figure 1.5: Overview of the Two-Dimensional Convolutional Neural Network Architecture.

CNN are designed to learn features using multiple building blocks, such as:

Convolution Layer

The name "convolution" indicates that the ANN employs a mathematical operation called "convolution" to extract features and reduce dimensionality. A convolution operation creates a feature map of the input by applying a kernel relatively smaller than the input that moves along all the input data by the stride value and performs an element-wise multiplication over the overlapping values of the kernel and the input segments (one segment at a time) with respect to the size of the filter, and finally summing up the obtained results. The process is the same when the data is 1D and 2D. Later, the feature map obtained is then fed to other layers to learn several other features.

Pooling Layer

In most cases, the convolution layer may be succeeded by a Pooling layer; it can be either an average-pooling or a max-pooling layer. The aim of these layers is to decrease the size of the convolved feature map to reduce the computational costs.

Average-pooling layer takes all of the features into consideration and calculates the average of the elements in the feature map in the pooling region. And max-pooling layer, on the contrary, only picks the most active features in a pooling regions.

Fully-connected Layer

Following the aforementioned layers, comes the fine-tuning stage where a flatten layer reshapes the obtained feature maps into a 1D array, which serves as the input to the FC layers also known as Dense layers that-are placed before the output layer. At this stage, the ANN is trained as would be trained a simple MLP [18].

1.5 Conclusion

AI has grown to be very popular and interdisciplinary field in today's world of scientific computing. Using AI algorithms and more precisely DANN we are allowed to make sense of large amount of different data-sets in different domains.

This chapter serves as guide for Deep Learning (DL) methods. The theories behind several DANN methods related to our research work are explained, offering a comprehensive guide for building and deploying several techniques.

Due to their ability to model complex relationships and patterns in data, DANN achieved SOTA performance on a wide range of tasks across various domains, such as forecasting and image classification, while tackling unimodal problems; they are undoubtedly worthy of investigation when the problems are extended to multi-modal ones.

Chapter 2

Multi-modal Learning

2.1 Introduction

As human beings, our experience to the world is inherently multi-modal. Our cognitive abilities are often associated with successful learning from multiple modalities simultaneously. Thus, in the context of human perception, We can think of a mode as a basic human sense. We can see objects smell odours, feel textures, taste flavours and hear sounds. Furthermore, when we employ multiple senses simultaneously to process one task, it would be more accurate to get the desired result, case in point, if we hear someone's voice and see that person, we would be able to identify him faster than if we hear the voice blindfolded. Consequently, it could be said that the multi-modality approach to perceiving the world is a human related ability.

So, in the context of AI, What is multi-modality? and how machines can simulate this process?

In this chapter we will answer this question in details providing types of different modalities in AI. We will also review the taxonomy of Multi-modal AI and more precisely Multi-modal ML.

2.2 Multi-modal Deep Learning

Following the working principle of our brain, directly or indirectly, drawing consolidated conclusions involves the interaction of different streams of information. Our brain is continuously processing and understanding an overlap of multiple modalities simultaneously and giving us information about the

environment around us. Hence, more recently AI begun to reverse the dominance of mono-modality in ML and it has been expanded to involve a variety of data modalities to solve more challenges, leading to a better understanding and analysis of the information and therefore achieving better accuracy.

Multi-modal DL is a ML sub-field that aims to train AI algorithms. The latter more often than not, rely on DANN to process and find relationships between different modalities to overcome the limited capabilities of the unimodal DANN.

We have all heard the saying that 'A picture is worth 1,000 words', which is why in certain situations the interplay between different modes of information has the potential to convey an idea more effectively and more quickly than a single mode. In general terms, a modality refers to the way in which a real-world concept is depicted, and hence described using an expressive mode. However, in AI, a modality or more explicitly a modality of information representation is a way of coding or representing information in some medium [19]. Multi-modality on the other hand or Multi-modal DL in AI is a vibrant multi-disciplinary research field that aims to design DL models in a way that enables them to learn through integrating multiple communicative modalities together that are involved to create meaning and generate a better result. It helps researchers to describe a problem using different aspects or viewpoints with complementary or supplementary information [20].

2.2.1 The Different Modalities in the Context of Artificial Intelligence

Each individual modality in AI uses a unique semiotic resources to create meaning [21]. To be specific, the term "Modality" refers to how things are experienced in terms of sensory inputs or how information is represented and communicated [20]

The term "modal" or "modality of data" is a reference but not limited to the data types that can be used by DL models like text, image, video, etc. To be more specific, another example of modalities from the same data type that are considered as different modalities are: Different languages, different image color scales, data and its corresponding metadata etc.

Each modality presents its own challenges and opportunities for AI, and systems can often be designed to handle multiple modalities at once for a more complete experience. A research problem is therefore characterized as

multi-modal when it includes two or more of such modalities.

From a modality perspective, all data can be grouped into two categories: structured and unstructured.

Structured Data

Also known as Quantitative Data in statistics, structured data is the most basic type of modalities that expresses information in the form of exact numbers. It refers to data that is neatly organized in a specific, well-defined format, it is often stored in relational databases or spreadsheets and can be easily searched and sorted.

It can be characterized by discrete or continuous data. Discrete data has distinct values, whereas continuous data are further grouped into interval and ratio data. This data type has meaning as a measurement such as electric load consumption per hour or temperature of the day, and what differentiates it from other data types, is its ability to carry out arithmetic operations.

Unstructured Data

Structured data is contrasted with unstructured data, which has no predefined format, it is the most abundant type of data available that includes data in the form of texts, images, and audio. While unstructured data can be more challenging to work with, it also contains richer and more detailed information that helps DL algorithms better analyze and gain insights from it. Here are some common unstructured modalities in the context of AI:

Visual modality

As the name indicates, visual modality refers to unstructured kind of modalities that is in the form of visual representation of something captured by cameras or scanners for instance, such as digital images represented by a 2D or 3D arrays of pixels containing intensities of particular color channels. It could be also a digital video. This type of data is designed to be easily interpreted by humans and can convey complex information more effectively than text or numerical data alone. In AI, visual data plays a crucial role in tasks such as image classification.

Computer vision is the corresponding technology for visual data, it helps AI systems make sense of complex information and enables humans to gain insights and make informed decisions. The visual modality has broad applications such as image recognition, object detection, and image segmentation.

Textual Modality

Textual or Natural Language modality is a critical component of AI systems, it refers to the kind of unstructured data that is in the form of plain text gathered via digital documents, scanners or cameras. Natural language processing is a sub-field of AI referring to the ability of AI systems to process, understand and transform the unstructured text in databases into normalized and structured data so that it would be possible to infer meaning from it using DL algorithms. Natural Language is a key aspect of many AI applications such as voice assistants, chatbots, and machine translation systems. It involves techniques such as language translation and sentiment analysis.

Auditory Modality

The auditory modality refers to the use of sound as a means of communication or expression that conveys information or emotion. Auditory modality is a common data type in AI that is used in several application such as speech recognition. It exists as digital information encoded as audio files, this digital information is broken down into thousands of fragments per time-factor where each fragment is stored as binary data so that it could be processed by machines to make AI and virtual assistants programs more user-friendly. Auditory data is collected from a variety of devices usually microphones, and it comes in many forms in the real world;

- Voice: Sound produced by humans as utterances or songs.
- Speech: A recording of a conversation between two or more speakers.
- Sound: Including music, it can be also a sound made by an animal, a car or an alarm for instance.

2.3 Multi-modal Machine Learning Taxonomy

Multi-modal ML taxonomy refers to the categorization of ML techniques that involve multiple modalities or forms of input data. Some common modalities include audio, visual and textual data. The taxonomy of multi-modal ML involves 5 fundamental challenges: Representation, Translation, Alignment, Fusion, and Co-learning.

The aforementioned aspects are shown as distinct, but in reality there is an overlap among them, it could be necessary to apply multiple techniques together to achieve several tasks. For instance, representation is required to interpret the heterogeneous data before fusion task and vice versa, we may need fusion to achieve a unified representation of the data. Likewise, representation is also required for the tasks related to translation and alignment and methods from all remaining four challenges are used to achieve the co-learning [20].

2.3.1 Multi-modal Representation

Representing and summarizing data from multiple modalities in a way that exploits the complementarity and redundancy is what we call multi-modal representation [22] Multi-modal representation refers to the encoding or representation of information from multiple modalities, such as audio, visual and text, into a single unified representation. These representations map multiple modes of information to a single mathematical space [23]. It can be used as input for DL algorithms, allowing them to learn from and make predictions based on several types of information.

Multi-modal representation is a key aspect of multi-modal ML, as it enables the integration and analysis of information from different sources, which can lead to improved performance and accuracy. There are various techniques for creating multi-modal representations, including DANN architectures that can handle multiple inputs.

In the past decade, many multi-modal representation learning algorithms have been proposed by different researchers, these algorithms can be roughly classified into two categories: Joint representation algorithms and coordinated representation algorithms [24].

Joint Multi-modal Representation

The main idea of joint representation involves the integration of the unimodal features together into the same representation shared space to learn a better representation of the data while preserving information from the given modalities. This means that the model learns to relate elements from one modality to elements of another modality. The fundamental premise is that there must exist a connection between the modalities, meaning that there must be a coherent semantic relationship between them. To make this more

concrete, one approach is to combine information from different modalities into a unified input. The joint input is a novel entity comprising numerous modalities, yet it is processed as a unified input [25].

In joint representation, data from all modalities is required at training and inference time which can potentially make it hard dealing with missing data [22].

Mathematically, the joint representation is expressed as Eq. (2.1) [22]:

$$x_m = f(x_1, \ldots, x_n) \tag{2.1}$$

The multi-modal representation x_m is computed using a function f, e.g. DANN that relies on multiple uni-modal representations $x_1 \ldots x_n$.

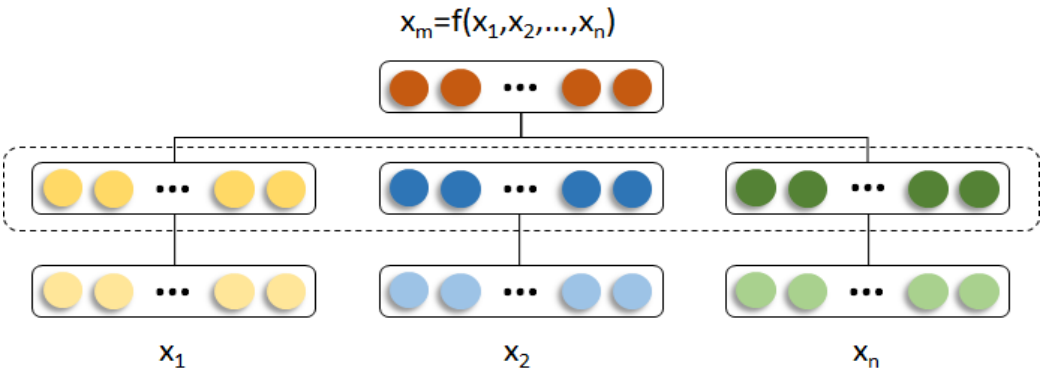


Figure 2.1: Joint Representation Learning Framework.

Coordinated Multi-modal Representation

In coordinated representation, each model will learn to project one modality into its own representation space. These uni-modal representations are coordinated through their interconnections separately [26], by applying a certain similarity constraint to bring the separated representations closer into a what we term coordinated space [27], where the learning objective is to preserve both inter-modality and intra-modality similarity structure. This constraint refers to a loss function identifying cross-modal similarity/correlation. Example constraints: minimize cosine similarity, maximize correlation.

Coordinated representations have a huge advantage over joint representations because when the modalities are very different fundamentally, this ap-

proach is considered more practical, due to the variety of modalities [25, 22].

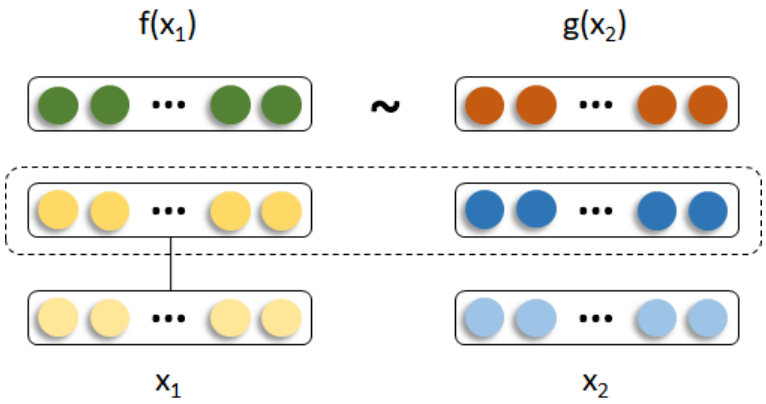


Figure 2.2: Coordinated Representation Learning Framework.

2.3.2 Multi-modal Alignment

Another multi-modal challenge is in the identification of cross-modal connections and interactions among elements from several modalities. Multi-modal Alignment refers to the process of synchronizing instances from at least two modalities by finding their links and correspondence. For instance, When analyzing speech and related gestures of a human subject, the objective is to synchronize uttered words with certain motions. Similarly, when provided with a caption and a picture, we must identify the regions of the image that correlate to the words in the caption [22].

2.3.3 Multi-modal Fusion

Multi-modal fusion is a paramount research direction in multi-modal learning; it refers to the process of joining information from multiple modalities, e.g. visual and text, to make a decision or to generate a more complete and robust representation of the input. The goal is to leverage the strengths of each modality to overcome the limitations of any single modality, and to provide better overall performance in a given task.

The interest in multi-modal fusion arises from three benefits it may provide. Firstly, learning from diverse modalities that observe the same phenomenon may result in more robust predictions. Secondly, leveraging mul-

tiple sources of information has the potential to capture complementary information — something that is not visible in individual modalities. Finally, when one of the modalities is missing, a multi-modal system can still operate.

Multi-modal fusion using DL approaches is classified into three main strategies: Early fusion, late fusion and hybrid fusion.

Early Fusion

Early Fusion is a type of multi-modal fusion that involves combining data from multiple modalities at an early stage of the processing pipeline, it allows the different modalities to influence each other by exploiting the correlation and interactions between low level features of each modality when the feature extraction processing takes place. Early Fusion can be performed on raw data directly, or on feature level, where instead of directly fusing raw modalities, feature-level fusion attempts to extract certain features from raw data through DL models before carrying out the data fusion [28].

Various feature learning sets show different characteristics of same pattern and combining those features retain active discriminant information while completely remove the redundant information [29].

Furthermore it has an advantage over late fusion, because it creates a single integrated representation that incorporates information from all modalities which only requires the training of a single model to make decisions. However, early fusion can also result in increased complexity and hence needs more sophisticated models.

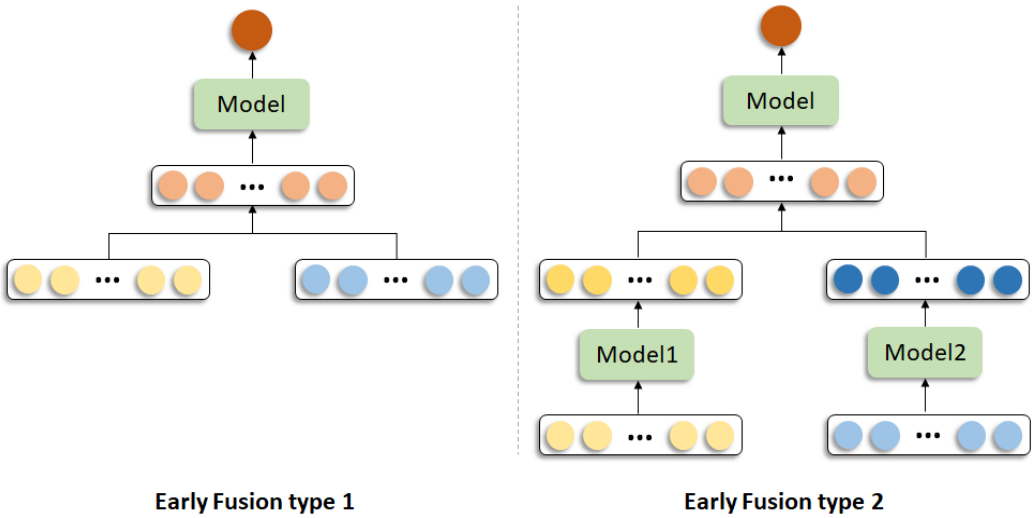


Figure 2.3: Early Fusion Framework.

Late Fusion

In contrast to early fusion strategy, late fusion is a type of multi-modal fusion that occurs at the decision-level, it allows the different modalities to be processed independently allowing the use of different models each of which is trained on a single modality, which can simplify the processing pipeline and reduce the computational complexity. The idea is to obtain independent predictions from each modality and then combine them to make a final decision by voting, finding maximum, or averaging [30].

On one hand, the advantage of late fusion lies in is the possibility to visualize the decision result of each modal based on a single modality separately. Although this strategy ignores the low-level interaction between modalities, it permits an easy model training with more simplicity and flexibility to make predictions when some modalities are missing. On the other hand, late fusion may not capture correlations and dependencies between modalities, which can lead to a poor performance compared to early fusion.

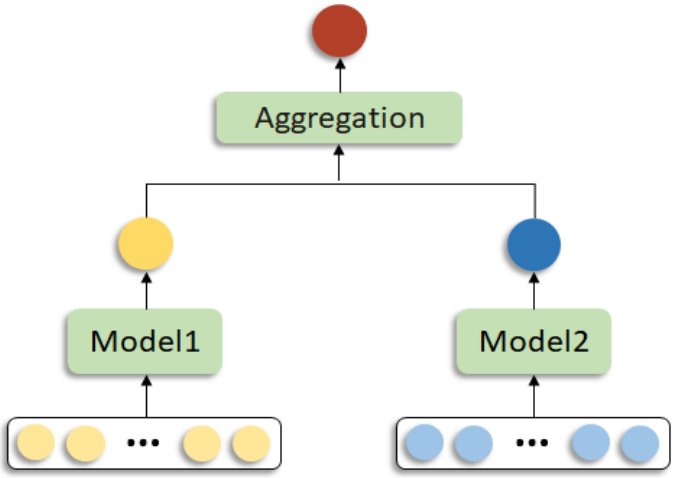


Figure 2.4: Late Fusion Framework.

Hybrid Fusion

Hybrid Fusion is a type of multi-modal fusion that attempts to exploit the advantages of both early and late fusion in a single framework [31]. The idea is to take advantage of the strengths of both strategies, while avoiding their limitations. In a hybrid fusion system, some modalities may be combined at an early stage, while others are combined at a later stage. The choice of when to perform fusion depends on the task and the modalities involved.

Depending on the number of the present modalities in the system, there is an infinite ways to merge modalities in a hybrid strategy, for instance we perform an early fusion on two modalities, and then the output of that multi-modal network is merged with another modality.

2.3.4 Multi-modal Translation

Unlike traditional translation that deals with text, in certain ML problems, the translation requires incorporating information from diverse sensory modalities to broaden its scope, such as translating image to a textual caption that describes the visual image or vice versa. Another example is to transform textual documents from one language to another considering the two languages as different modalities. Thus, Multi-modal translation points out

to the Process of converting the representation of one source modality into that of another target modality. It involves learning a generative process to produce raw modalities that reflect cross-modal interactions, structure, and coherence [26]. The translation task can be supported by auxiliary modalities such as text, audio etc.

This translation is categorized into two types: Example based and generative. This distinction is similar to the one between non-parametric and parametric ML approaches respectively [22].

In [26] they distinguished three categories based on the information contained within input and output modalities. (1) Summarization compresses data through knowledge abstraction into a condensed version retaining the most important information within the original content. (2) Translation maps data from one modality to another while respecting cross-modal interactions. Finally, (3) creation aims to generate novel and coherent high-dimensional multi-modal data from small initial examples or latent conditional variables.

Example Based Models

Example-based multi-modal translation algorithms are restricted by their training data, they involve using specific examples as a basis for translating content across different modalities. This approach leverages explicit examples to guide the translation process. The goal is to create a system that can learn from specific instances and generate accurate translations for diverse modalities.

when translating one modality to another, there is two types of such algorithms: retrieval based, and combination based. Retrieval-based models are the simplest form of multi-modal translation, they rely on simply using the retrieved translation without modifying it, once the closest sample in the dictionary is found, it is therefore used as the result. While combination-based models rely on more complex rules to create translations based on a number of retrieved instances. Instead of just retrieving examples from the dictionary, they combine them in a meaningful way to construct a better translation [26].

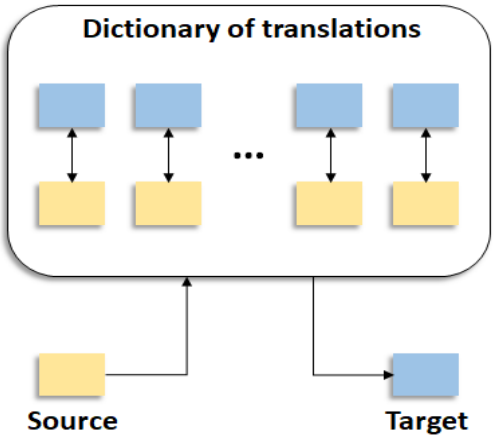


Figure 2.5: Example Based Translation Framework.

Generative Based Models

In contrast to example-based models, generative models are able to perform a multi-modal translation on a uni-modal source. In this case, we can train generative multi-modal translation models that learn to decode samples from a vector space into an output of a different modality. It is a more challenging task as it requires the ability to both understand the source modality and to generate the target modality.

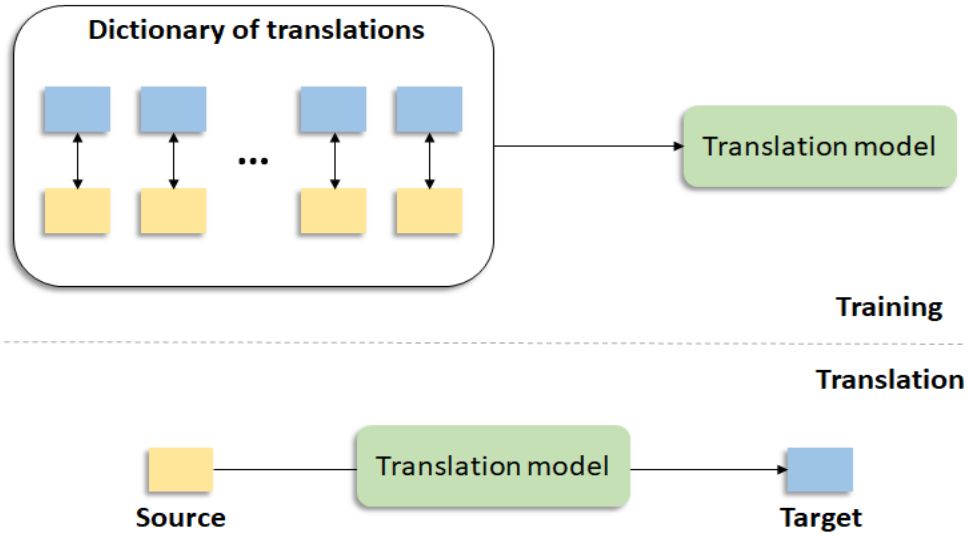


Figure 2.6: Generative Based Translation Framework.

2.3.5 Multi-modal Collaborative Learning

In a standard multi-modal task, it is commonly assumed that all the modalities are available during training and testing. A key objective of multi-modal co-learning is to work in real-life conditions where one or more modalities are scarce during training. Those may be noisy, lack of annotations, have unreliable labels, have limited resources or may not be available at all times [20].

Multi-modal Co-learning aims to transfer information learned through one or more modalities to tasks involving another, where one modality acts as a supporting modality and aids another modality with the transfer of knowledge between modalities during training. This supporting modality is normally not present at inference time. The same is applicable for more than two modalities[20].

2.4 Conclusion

In this chapter, we presented the main concepts of multi-modality by detailing the different modalities in the context of AI as well as providing a detailed review about the taxonomy of multi-modal ML in the form of challenges and

possible applications such as multi-modal representation, translation, alignment, co-learning and fusion.

In conclusion, multi-modal learning represents a paradigm shift in the field of ML and AI. Its ability to integrate and process data from multiple modalities has paved the way for the development of advanced AI systems capable of comprehending, analyzing, and interpreting information in a more nuanced manner, leading to improved decision-making and task performance.

Chapter 3

The Intersection of Multi-modal Deep Learning and Different Domains and Applications

3.1 Introduction

Multi-modal AI allows for a more intelligent processing of multiple data modalities simultaneously, which is one factor motivating the development of multi-modal DL. Multi-modal DL is used to denote a highly multidisciplinary research field and set of DL methods to extract high level information from low-level multi-modal data. Various multi-modal methods and applications are designed using multi-modal DL techniques and their use is extensive and encompasses a wide range of potential applications in different domains. In recent years, research on multi-modal tasks has gradually emerged. In the present chapter, we grouped the existing methods and applications with relevance to multiple research areas in order to highlight recent advances and trends.

3.2 Multi-modal Applications

Various methods and applications are designed using multi-modal DL techniques. In this section, we present the most popular multi-modal applications

with different related works found in the literature. The different authors' contributions are grouped with relevance to the multi-modal application domain.

3.2.1 Multi-modal Image Description

Multi-modal image description or image captioning is at the intersection of computer vision and natural language processing. It refers to the task of generating textual description for a given image. SOTA methods involve processing information from both the visual content of an image and any accompanying textual information to create a coherent and descriptive caption. Multi-modal image description has seen remarkable progress over the years, owing it to the advancements in computer vision and natural language processing techniques.

In this work [32], an image captioning model named (AoANet) was proposed as a baseline model to leverage the text detected in the input image. In addition, a pointer-generator mechanism has been used to copy the detected text to the caption in order to reproduce tokens accurately if needed.

Web data sets come with a lot of noise, and existing data filtering methods are effective at reducing noise from data sets, but they also come at the expense of the diversity of the original training data. The effectiveness of synthetic captions in improving caption quality of these data-sets for multi-modal training has been demonstrated in [33].

In this paper [34], a novel multi-modal transformer [35] framework for image captioning was proposed. It consists of an image encoder that generates visual representations via deep self-attention learning and a caption decoder to transform the encoder's visual features into textual captions.

Another work on multi-modal fusion for image captioning has been proposed in [36]. The proposed method combines image high-level attributes and sentence representation using ATT_{ssd} and CNN_m based on Flickr8k, Flickr30k and MSCOCO data-sets.

In this paper [37], the limitations of using pre-trained frozen object detectors such as CLIP are addressed while tackling the image captioning task. They proposed to add an auxiliary branch in the graphical model, leveraging advances in large pre-scaled multi-modal models to retrieve contextual attribute and relationship descriptions.

Title	Approach	data-set	Modalities	Ref
Multi-Modal Image Captioning for the Visually Impaired	AoANet	VizWiz	text, image	[32]
Improving Multi-modal data-sets with Image Captioning	CLIP, BLIP, BLIP2	Flickr30K, MS-COCO, LAION-COCO	text, image	[33]
Multimodal Transformer With Multi-View Visual Representation for Image Captioning	Multimodal Transformer	MSCOCO	text, image	[34]
A multimodal fusion approach for image captioning	CNN, GRU, LSTM	Flickr8k, Flickr30k, MSCOCO	text, image	[36]
Beyond a Pre-Trained Object Detector: Cross-Modal Textual and Visual Context for Image Captioning	CLIP	Visual Genome, MS-COCO	text, image	[37]

Table 3.1: Summary of Multi-modal Image Description Contributions.

3.2.2 Multi-modal Image Retrieval

Multi-modal image retrieval is a task in which information is retrieved from a data-set containing images using queries that encompass multiple modalities, such as text, images, audio, or any combination thereof, typically, combining visual and textual information. The goal is to find relevant images that

match both the visual content and associated query. Many researchers have made significant contributions in this field and have proposed many methods for multi-modal image retrieval.

This contribution [38] presents a novel image-text composition module based on additive attention that can be seamlessly plugged into DANN to tackle the challenging task of image search. Given a source image and text feedback, the goal is to retrieve the target images that resemble the source yet satisfy the given text by composing a multi-modal image-text query.

While in [39], authors present a multi-modal framework based on vision-language fusion named (EnVLF). It consists of two uni-modal (vision and language) encoders and a multi-modal encoder.

Learned Sparse Retrieval (LSR) is a group of neural methods designed to encode queries and documents into sparse lexical vectors. These vectors can be efficiently indexed and retrieved using an inverted index. In this work [40], they explored the application of learned sparse retrieval for the image suggestion task to support multimedia content creation.

This work [41] contributes to the development of an automated annotation model that retrieves visually similar images from online multimedia streams based on RL algorithm and a multi-modal active learning that uses a convolutional recurrent neural network for automatic annotation of labels using visually similar contents or features like edges, color, texture, shape, and spatial information.

A bi-directional training scheme is proposed in [42] for composed image retrieval that exploits information from the mapping of the target image-text pair to the reference image. To tackle the challenge of inferring the reversed semantics of the text with the absence of additional annotations, they leverage the text encoder and prepend learnable tokens to the text input.

Title	**Approach**	**data-set**	**Modalities**	**Ref**
Fashion Image Retrieval with Text Feedback by Additive Attention Compositional Learning	Additive Attention Compositional Learning	FashionIQ, Fashion200k, and Shopping100k	text, image	[38]

Title	Approach	data-set	Modalities	Ref
An End-to-End Framework Based on Vision-Language Fusion for Remote Sensing Cross-Modal Text-Image Retrieval	Encoder, Vision transformer	RSICD and RSITMD	text, image	[39]
Multimodal Learned Sparse Retrieval for Image Suggestion	Learned Sparse Retrieval model	AToMiC	text, image	[40]
Multi-modal active learning with deep reinforcement learning for target feature extraction in multi-media image processing applications	Multi-modal Active Learning, Reinforcement Learning	-	text, image	[41]
Bi-directional Training for Composed Image Retrieval via Text Prompt Learning	BLIP, bi-directional training scheme	Fashion-IQ, CIRR	text, image	[42]

Table 3.2: Summary of Multi-modal Image Retrieval Contributions.

3.2.3 Multi-modal Visual Question Answering

Visual question answering is another intersecting research area of computer vision and natural language processing that requires a system to do much more than task specific algorithms. It is a multi-modal task that uses natural language to ask and answer open-ended questions based on images [43].

Various methods were proposed for multi-modal visual question answering in the literature.

In [44], a multi-modal co-Attention relation network that combines co-attention and visual object relation reasoning was proposed. It uses the co-attention module to learn the textual features and object-level visual representations that are more critical for predicting the correct answers, and further utilizes the visual relation reasoning module to model the visual representations at relation-level.

This paper [45] proposes a novel and efficient multi-modal adaptive gated mechanism model to obtain accurate modality feature information. It adds an adaptive gate mechanism to the intra- and inter-modality learning and the modal fusion process.

Another visual question answering method that performs bidirectional fusion between structured and unstructured multi-modal knowledge to obtain unified knowledge representations is illustrated in [46].

In this paper [47], they introduced an approach that leverages multi-modal fusion and cross-modal contrastive learning in video question answering. First, a gated multi-modal fusion network combines different modalities such as visual and speech modalities based on their relevance to the question to enrich the representations of video and improve the accuracy. Second, they introduced cross-modal contrastive learning to increase the similarity between positive example pairs.

In [48] NuScenes-QA data-set for visual question answering in the context of autonomous driving was introduced, aiming to answer natural language questions based on street-view clues. The raw visual data includes images and point clouds captured by camera and LiDAR.

Title	**Approach**	**data-set**	**Modalities**	**Ref**
Multi-modal co-attention relation networks for visual question answering	Multi-Modal Co-Attention Relation Network	VQA	text, image	[44]
Multi-modal adaptive gated mechanism for visual question answering	multimodal adaptive gated mechanism model	VQA and GQA	text, image	[45]

Title	Approach	data-set	Modalities	Ref
VQA-GNN: Reasoning with Multimodal Knowledge via Graph Neural Networks for Visual Question Answering	Graph neural network bidirectional fusion method	VCR and GQA	text, image	[46]
Gated Multi-modal Fusion with Cross-modal Contrastive Learning for Video Question Answering	Gated Multi-modal Fusion and Crossmodal Contrastive Learning	AVQA and Music-AVQA	text, video, audio	[47]
NuScenes-QA: A Multi-Modal Visual Question Answering Benchmark for Autonomous Driving Scenario	camera and LiDAR	NuScenes-QA	text, image, point clouds	[48]

Table 3.3: Summary of Multi-modal Visual Question Answering Contributions.

3.2.4 Multi-modal Emotion Recognition

Multi-modal emotion recognition allows the identification and understanding of human emotional states by combining information from different modalities, including but not limited to textual information, speech, and facial expressions as emotions are often expressed through a combination of voice tone, facial expressions and body language [49].

Multi-modal emotion recognition has never failed to attract researchers'

interest.

In [50], researchers proposed a model consisting of uni-modal transformer modules that learn representations for each modality separately, and a multi-modal transformer module that fuses all modalities to perform the emotion recognition task.

[51] is another work presenting a new way of multi-modal emotion identification. This method designs a multi-level CNN based on facial expression modality and a stacked bidirectional LSTM model based on electroencephalograhpy (EEG) modality. At the decision level, D-S evidence theory [52, 53, 54] is used to fuse the emotion identification results.

This work [55] presents multi-label emotion recognition model based on adversarial temporal masking strategy, it consists of encoding other modalities by masking the most emotion-related temporal units e.g. words for text or frames for video of the informative modality. Additionally, an adversarial parameter perturbation strategy was proposed to enhance the model's generalization by adding the adversarial perturbation to the parameters of the model.

This work [56] focuses on developing a hierarchical cross-attention model approach to recognize emotional state using a combination of recurrent and co-attention neural network models. Audio data is processed through a learnable Wav2vec approach and, the textual data is represented using a bidirectional encoder representations from transformers BERT model.

Another study that focuses on developing a multi-modal emotion recognition system based on audio and text modalities known as MM-EMOR was proposed in [57]. The use of mel spectrogram features, chromagram features, and the mobilenet CNN for processing audio data are central to the operation of this system, while an attention-based roberta model caters to the text modality.

Title	**Approach**	**data-set**	**Modalities**	**Ref**
Layer-wise Fusion with Modality Independence Modeling for Multi-modal Emotion Recognition	multi-modal transformer	CHERMA	text, video, audio	[50]

Title	Approach	data-set	Modalities	Ref
Multi-modal emotion identification fusing facial expression and EEG	convolutional neural network, stacked bidirectional LSTM and D-S evidence theory	CK+ and Fer2013	image, EEG	[51]
Learning Robust Multi-Modal Representation for Multi-Label Emotion Recognition via Adversarial Masking and Perturbation	adversarial temporal masking strategy and adversarial parameter perturbation strateg	CMU-MOSEI and NEMu	text, video	[55]
HCAM – Hierarchical Cross Attention Model for Multi-modal Emotion Recognition	wav2vec, BERT and bi-directional recurrent neural network	IEMOCAP, MELD and CMU-MOSI	text, audio	[56]
MM-EMOR: Multi-Modal Emotion Recognition of Social Media Using Concatenated DL Networks	Mobilenet, attention-based Roberta model	Tweeteval, MELD, IEMOCAP	text, audio	[57]

Table 3.4: Summary of Multi-modal Emotion Recognition Contributions.

3.2.5 Multi-modal Speech Synthesis

Multi-modal speech synthesis presents a challenging yet promising area of research. It deals with automatic speech generation through capturing multi-modal data from real speakers synchronously including speech, natural language and facial expressions, and then exploiting them for automatic speech generation. In other words, an unseen speaker's voice can be built using another types of data modalities.

A big interest was shown to multi-modal speech synthesis in the literature.

In this paper [58] a multi-scale multi-modal conversational text-to-speech system was proposed which is a multi-modal translation task that aims to synthesize speech.

Another powerful and efficient image to speech captioning model was built in [59] based on pre-trained vision language model.

Based on encoder-decoder, [60] proposed a tri-modal translation model that translates between arbitrary modalities spanning speech, image, and text, treating different modalities as different languages

In this paper [61], they proposed AudioGPT a multi-modal system providing large language models with complex audio information processing modules as well as input/output interface (ASR, TTS) to support spoken dialogue.

This work [62] focuses on implementing a multi-modal style transfer task in text to speech framework named MM-TTS based on a constructed data-set called MEAD-TTS. It can use any modality to control the style of the generated speech, including reference speech, emotional facial images, and text descriptions. An aligned multi-modal prompt encoder that embeds different modalities into a unified style space, supporting style transfer for different modalities was firstly proposed. Additionally, a new adaptive style transfer method named Style Adaptive Convolutions (SAConv) to achieve a better style representation was proposed. Furthermore, they designed a Rectified Flow based Refiner to solve the problem of oversmoothing Mel-spectrogram and generate audio of higher fidelity.

Title	Approach	data-set	Modalities	Ref
M2-CTTS: End-to-End Multi-Scale Multi-Modal Conversational Text-to-Speech Synthesis	BERT, Wav2vec	DailyTalk	text, audio	[58]
Towards Practical and Efficient Image-to-Speech Captioning with Vision-Language Pre-training and Multi-modal Tokens	Vision Transformer	Flickr8kAudio and SpokenCOCO	text, image, audio	[59]
TMT: Tri-Modal Translation between Speech, Image, and Text by Processing Different Modalities as Different Languages	multi-modal encoder-decoder	CC3M, CC12M, COCO, SpokenCOCO, Flickr8k, Flickr8kAudio	text, image, audio	[60]
AudioGPT: Understanding and Generating Speech, Music, Sound, and Talking Head	gpt-3.5-turbo	LJSpeech, AudioCaption	text, audio	[61]
MM-TTS: Multi-Modal Prompt Based Style Transfer for Expressive Text-to-Speech Synthesis	CLIP, Style Adaptive Convolutions	MEAD-TTS	text, image, audio	[62]

Title	Approach	data-set	Modalities	Ref

Table 3.5: Summary of Multi-modal Speech Synthesis Contributions.

3.2.6 Multi-modal Event Detection

Social media user-generated content provides a valuable source of multi-modal data for researchers to detect real-world events. The task of event detection refers to the classification of content into event and non-event as well as the classification of event types e.g. sport, music, breaking news etc.

Online media sharing allows users to effortlessly share events, activities, and thoughts at any time, making multi-modal data-sets available for research purposes.

A temporal multi-modal graph learning method was proposed in [63] for acoustic event classification. Such temporal information was modeled via graph learning techniques, where one temporal graph was constructed for each acoustic event, dividing its audio data and video data into multiple segments. Each segment can be considered as a node, and the temporal relationships between nodes can be considered as timestamps on their edges.

Event extraction aims to extract information from diverse modalities. However, the different modalities are often misaligned at the event level. To address this issue, they firstly introduced a new data-set called TVEE, which consists of sentence-video pairs describing the same events. They also constructed a new cross-modal contrastive learning model for event extraction [64].

This study [65] presents a transformer based architecture for detecting apnea events from commonly collected sleep signals derived from PSG and patients' electronic health records.

This study [66] proposes Human Cognition-based Consistency Inference Networks to comprehensively explore consistent and inconsistent semantics for multi-modal fake news detection. Specifically, a cross-modal alignment layer was designed to learn consistent semantics between textual and visual information within the multi-modal news, and then the comment clue discovery layer is devoted to ascertaining the most-concerned semantics by audiences between comments. Then, collaborative inference layer was developed to drive news consistent semantics and the most-concerned semantics to

reason and discover consistent and inconsistent information between them.

Misfire detection in diesel engines can potentially extend their operational lifespan, ensuring safety, and yield economic and societal benefits. In this paper [67], a multi-modal transformer-based CNN was proposed under significant environmental noise and diverse working conditions. Vibration signals were collected from engine cylinder heads at different speeds and extracted both one-dimensional amplitude vector features and two dimensional image properties. These features were input into the multi-modal feature extraction network and subsequently processed by the cross-channel feature fusion network, the latter incorporated a dual-dimension spatial and channel attention mechanism to suppress vibration noise interference.

Title	**Approach**	**data-set**	**Modalities**	**Ref**
TMac: Temporal Multi-Modal Graph Learning for Acoustic Event Classification	Multi-modal graph learning	AudioSet	audio, video	[63]
Cross-Modal Contrastive Learning for Event Extraction	BERT, Conditional Random Field, Transformer	TVEE, VM2E2	text, video	[64]
Bringing At-home Pediatric Sleep Apnea Testing Closer to Reality: A Multimodal Transformer Approach	transformer	Nationwide Children's Hospital (NCH) Sleep Data Bank and Childhood Adenotonsillectomy Trial (CHAT)	ECG and SpO2 signals	[65]

Title	Approach	data-set	Modalities	Ref
Human Cognition-Based Consistency Inference Networks for Multi-Modal Fake News Detection	BERT, ResNet-152	Weibo, Twitter, and PHEME	text, image	[66]
MITDCNN: A multi-modal input Transformer-based deep convolutional neural network for misfire signal detection in high-noise diesel engines	transformer, convolutional neural network	-	image, amplitude	[67]

Table 3.6: Summary of Multi-modal Event Detection Contributions.

3.2.7 Multi-modal Action Recognition

Multi-modal action recognition also referred to as multi-modal activity recognition is the process of detecting and identifying physical actions or behaviors from multiple modalities or sources of information. In smart homes for instance, sensor data allows for energy saving by detecting when a person enters or leaves a room to adjust the lighting or temperature accordingly. It can also include a wide range of applications such as human-computer interaction, video surveillance, and robotics.

One of the most broadly studied multi-modal application is multi-modal Action Recognition.

Some contributions propose multi-modal transformer networks for action recognition, In [68], a transformer was used to detect actions in untrimmed videos. While in [73], an audio-visual temporal action detection transformer was implemented to fuse audio and visual modalities in an end-to-end fashion.

To leverage complementary information across modalities, authors in [69]

introduced a recurrent unit called multi-modal contextualization unit, which is a core component of the model. It temporally encodes a sequence of one modality with action content features of other modalities to exploit global action content and also to supplement complementary information of other modalities.

This research [70] focuses on distinguishing the walking activity defined as a sequence of repeated leg-swing actions from repetitive leg-swing activities in sitting state. For this purpose, a heterogeneous sensor system is implemented based on CNN-1D, that acquires novel multi-modal data from low-cost leg-worn IMU sensors (m-module) and finger-tip based pulse sensors (p-module).

A graph neural network approach was proposed in this study [71] to directly recognize three human activities (eating, working, and reading) at home using vision and speech teaching data. In the proposed method, an activity classification was learned from a 3D detected object corresponding to the human position. Next, human utterances were used to label the activity from the collected human and object 3D positions.

To tackle the multi-modal action recognition task, this contribution was proposed taking into consideration circumstances where some modalities are not available at an inference time. A transformer-based fusion was proposed with a simple modular network, which learns missing modality predictive coding by randomly dropping modality features and tries to reconstruct them with the remaining modality features. Coupling these good practices, a model robust to modality missing was built [72].

Title	**Approach**	**data-set**	**Modalities**	**Ref**
A Multi-Modal Transformer network for action detection	transformer	THUMOS14 and ActivityNet	video	[68]
AV-TAD: Audio-Visual Temporal Action Detection With Transformer	Transformer	THUMOS14	audio, video	[73]

Title	Approach	data-set	Modalities	Ref
Modality Mixer for Multi-Modal Action Recognition	recurrent unit	NTU RGB+D 60, NTU RGB+D 120, and NWUCLA	image, depth information	[69]
A deep-CNN based low-cost, multi-modal sensing system for efficient walking activity identification	1-D Deep Convolutional Neural Network	-	leg-worn IMU sensor and a wrist-worn pulse sensor	[70]
An efficient activity recognition for homecare robots from multi-modal communication data-set	graph neural network	-	video	[71]
Towards Good Practices for Missing Modality Robust Action Recognition	ResNet34, transformer	NTU60, NTU120, NWUCLA, UWA3DII	video	[72]

Table 3.7: Summary of Multi-modal Action Recognition Contributions.

3.2.8 Other Multi-Modal Applications

Multi-modal application may involve training DANN on multiple modalities to tackle many other problems such as multi-modal prediction. The latter encompasses both classification and forecasting tasks. Classification aims at predicting the future class while forecasting aims at predicting the future

value [74]. We may combine text and video modalities to accurately classify a movie into Drama, Horror, Comedy, or Romance for instance.

The popularity of multi-modal prediction arises from its potential benefit that has been proved in many research papers in the literature.

In [75], based on transfer learning, a combination of image of the chest X-ray/CT scan and the clinical notes provided with the scan was used to identify COVID-19 cases.

Wind speed, wind direction, and air density modalities were fused to obtain a unified representation in this paper [76]. Then, a stacking DL model was proposed, which contains the bidirectional gated recurrent unit (BGRU) and leaky echo state network (LESN). The final wind power forecasting results are acquired by a meta-learning operator.

To enhance machine understanding of text-intensive images, KOSMOS-2.5 a multi-modal transformer architecture pre-trained on large-scale text-intensive images was proposed in [77]. The proposed model excels in two tasks: (1) generating spatially aware text blocks, where each block of text is assigned its spatial coordinates within the image, and (2) producing structured text output that captures styles and structures into the markdown format.

As the agriculture technologies continues to develop, this study [78] aimed to implement a multi-modal DL model based on CNN integrating UAV-based multi-spectral imagery and weather data for a precise rice yield predictions.

Pedestrian detection assists video surveillance in crowd monitoring, people counting, and event detection. This study develops a multi-modal pedestrian detection and classification using hybrid meta-heuristic optimization with DL (IPDC-HMODL) algorithm. It follows a three stage process namely multi-modal object detection, pedestrian classification, and parameter tuning [79].

Title	**Approach**	**data-set**	**Modalities**	**Ref**
Multi-modal image classification of COVID-19 cases using computed tomography and X-rays scans	VGG16, ResNet50, InceptionResNetV2 and MobileNetV2	-	text, image	[75]

Title	Approach	data-set	Modalities	Ref
Multi-modal multi-step wind power forecasting based on stacking DL model	bidirectional gated recurrent unit and leaky echo state network	-	numerical data	[76]
KOSMOS-2.5: A Multimodal Literate Model	Transformer	-	text, image	[77]
Multimodal DL for Rice Yield Prediction Using UAV-Based Multispectral images and Weather Data	Convolutional Neural Network	-	UAV multispectral images and weather data	[78]
Intelligent multimodal pedestrian detection using hybrid metaheuristic optimization with DL model	YOLO-v5, RetinaNet, kernel extreme learning machine	UCSD	video	[79]

Table 3.8: Summary of Other Multi-Modal Applications Contributions.

3.3 Conclusion

Multi-modal DL is a highly multidisciplinary research field and ongoing research continues to explore more advanced models and techniques to enhance multi-modal tasks' solving in various domains.

In this chapter we attempted to highlight the different multi-modal DL applications that process and link information using various modalities such as image, audio, video, text, physiological signals etc. We focused addition-

ally on an up-to-date review of numerous multi-modal advancements and trends where different approaches have been proposed.

Chapter 4

State of the Art of deep learning for Electrical Load Forecasting

4.1 Introduction

Load forecasting is divided into three categories based on the time horizon of planning techniques: STLF, which is essential for power suppliers to operate power plants, MTLF, which is used for resource planning, and LTLF, which is used by decision-makers from industry and other interested stakeholders to plan and build power supply and transmission facilities [80, 2].

Load forecasting involves estimating future load demand that may have commercial and technical implications if not done correctly. Power and energy providers use it in managing production and consumption fluctuations to reduce the risk of power shortages and overloads. It also improves efficiency and revenue for generating and distribution companies, ensuring sustainability and balance between production and consumption over time.

Time series load forecasting is a challenging and worthwhile task that has been the subject of a lot of studies in the literature, using DANN architectures as they offer a lot of promise for a reliable and precise load forecasting.

In this chapter, we aim to review the recent studies presented in the field of time series load forecasting based on different DANN architectures.

4.2 Load Forecasting based on Convolutional Neural Networks

A multitask CNN is suggested for predicting household energy consumption in [81]. The model comprises a forecasting branch utilizing multi-scale dilated convolutions to forecast household load based on the previous 24 hours of load data, and a household profile branch using a Deep Convolutional Auto-encoder to encode historical load profiles and capture individual behavior uncertainty. Finally, both branches are merged at the end of the network using a fully connected layer.

The paper [82] introduces a method for STLF of the Bangladesh power system, utilizing an Encoder-Decoder architecture combining CNN and LSTM. The output of the CNN block, in the form of a flattened vector, serves as the input for the decoding process in the LSTM unit, followed by a dense layer to generate the output. Through simulations, it is evident that the proposed model can effectively handle long sequence time series data and accurately predict future load demand over a significant time span, while also achieving the lowest values of MAE, RMSE, and MAPE in comparison to the LSTM, RBFN, and XGboost models.

In their paper, the authors of [83] introduce a load forecasting model known as a Dilated Convolutional Dense Network. This model effectively captures local trends and seasonal patterns in time series data to make predictions for the future. They have trained the model using various time series data with different frequencies and tested it on five different datasets: CIF-2016, small and medium enterprise buildings, residential buildings, smart meter London dataset, and Turkish electricity data. The results indicate improved performance in four out of the five datasets used.

The authors in [84] designed a DL model by utilizing CNN layers arranged in a pyramidal structure. Initially, they categorized a subset of energy consumers from the Smart Grid Smart City (SGSC) database into clusters using the DBSCAN method. The CNN layers were employed to extract features from the historical load of each cluster. These extracted features were then amalgamated to create training databases for each cluster. The input feature vector comprises the energy usage from previous time instances and details about the hour of the day and the day of the week, which are presented in the form of one-hot encoding.

The paper initially employs a feature selection algorithm based on Ran-

dom Forest and then introduces a Hybrid Neural Network STLF algorithm that relies on multi-model fusion. The primary architecture is built on CNN and Bi-GRU. Input data is acquired by utilizing long sliding time windows with varying steps, followed by the separate training of multiple CNN-Bi-GRU models. The final load value for forecasting is obtained by averaging the forecasting results of multiple models. The datasets are sourced from a region in New Zealand and another in Zhejiang, China. When compared to other DL models, this model demonstrated superior performance in terms of MAPE and RMSE [85].

The study in [86] introduces a load forecasting CNN based on ResNet. It suggests using a first branch with Dilated Convolution for forecasting to extract features from historical load sequences, while the other branch is dedicated to learning external factors. The output vectors from both branches are then combined using a fully connected network to produce the final output. This approach significantly improves forecasting accuracy without increasing parameters and operations. The model's performance has been demonstrated to outperform various models such as GRU and LSTM in single-step and 24-step building load forecasting applications.

The authors introduced a new ensemble model for deep learning and data processing in their publication [87]. The study consisted of two main phases. In the data processing phase, the raw load data was analyzed using the auto-correlation function and the input for the model was determined. The electricity load data was decomposed into stable modes called intrinsic mode functions using VMD. Based on the ACF-determined time distribution and time lag, the model's input was transformed into a $24 \times 7 \times 8$ matrix denoted as M, representing 24 hours, 7 days, and 8 IMFs. For load forecasting, CNN-2D was utilized to extract features from matrix M. The improved reshaped layer was used to reorder the extracted features based on time. A temporal convolutional network was then employed to learn the reshaped time series features and combined with a fully connected layer for prediction. The model's performance was evaluated in the Eastern Electricity Market of Texas by comparing it with VMD-CNN and other models.

The methodology presented in reference [88] uses statistical analysis and involves preprocessing time series data before tackling the issue of day-ahead load forecasting. It utilizes statistical learning and appropriate data transformation to convert the data into an image-like format, enabling the use of CNN based models that take advantage of stationarity and time locality. While LSTM's superiority and efficiency are widely recognized, the study

demonstrated that CNN based models offer a viable alternative for STLF problems compared to LSTM algorithms, leveraging the temporal locality of load time series similar to how image processing exploits spatial locality. CNN appear to achieve higher accuracy in cases of limited historical data, as indicated by the MAE and RMSE, showing that the proposed CNN algorithm's performance was enhanced compared to the LSTM model.

The inputs of the deep learning model in [89] consist of multidimensional information comprising numerical and textual data. The textual data is encoded using the one-hot method. To enhance feature extraction, a modified inception structure is incorporated into the framework composed of CNN and LSTM, and adaptive residual connection is introduced to address gradient diffusion issues as the model's layers increase. Finally, the comparison results and improvements are presented after the textual data is integrated and the model structure is modified. Particularly under heavy load, the forecasting error is reduced, which is advantageous for preventing transformer overload in the distribution grid.

The authors of this study [90] introduced a CNN-based model aimed at addressing the STLF challenge by predicting one day ahead. The model takes as input the following data: Hourly electrical loads, daily minimum, maximum, and average temperatures, as well as date-related details such as day of the week, season, and holiday indicators. To validate the model, real public data from the Romanian power system was utilized. Results indicated that the proposed approach demonstrated superior potential in accurate forecasting and strong generalization abilities when compared to the forecasting outcomes of the Romanian TSO, with the MAPE serving as the evaluation metric.

An approach known as TCN has been proposed for forecasting residential electricity usage. This method has the capability to maintain load data for longer durations and simultaneously process load information. It is designed to model the electricity consumption of household appliances as well as the overall electricity usage of residential structures, utilizing the AMPds2 smart meter dataset. Experimental results indicate that the TCN prediction framework is able to effectively monitor the variations in residential electricity consumption [91].

The approach described in [92] utilizes the Pearson correlation coefficient. An experiment on residential load forecast involves the use of a CNN model. Initially, 100 curves of household electricity are segregated into N clusters. Using the PSK means method they propose, they amalgamate multiple clus-

ters into a new set containing richer information and features, which serves as the input for the CNN. The findings suggest that the new method provides more precise forecasting data compared to traditional methods.

The method proposed in this paper [93] introduces a new approach to probabilistic load forecasting using CNN. To address the issue of insufficient probability information, the Load Range Discretization method is suggested to create discrete Load Probability Distributions for the training samples. Subsequently, CNN is trained to directly generate the LPD for the forecasting samples. In the case study, it is observed that the majority of real loads fall within the prediction intervals of 70–90% confidence levels, indicating the effectiveness of the proposed method. When considering the same conditions of prediction interval coverage probability, the Mean Width of Prediction Interval is notably smaller compared to ECNN, QCNN, QRA, EWNN, QELM, EDNN, and ELSTM.

In [94], a two-part structure is suggested for examining the STLF problem, involving feature processing and improved CNN classifiers. A combined two-stage model is utilized to process n-dimensional time sequence data for selecting crucial features. Furthermore, t-Stochastic Neighbourhood Embedding is employed for feature extraction to improve CNN accuracy and speed by reducing redundancy. Additionally, the Grid Search Algorithm is used to automatically determine the appropriate super parameters for ECNN in order to enhance classification performance. The numerical findings validate that the proposed framework demonstrates improved accuracy compared to the standard CNN.

Table 4.1 lists the papers that tackle the load forecasting problem using CNN.

Case study	**Contribution**	**Approach**	**Horizon**	**Ref**
Irish load profile: Smart Metering Electricity Customer Behavior Trails (CBTs)	Multitask Deep ANN for load forecasting at household level	Multiscale Dilated CNN, Deep Convolutional AE	Very short-term	[81]
Bangladesh Power System	Electrical load forecasting	CNN, LSTM	Short-term	[82]

Case study	Contribution	Approach	Horizon	Ref
CIF2016 Energy Forecasting Competition, Small and medium enterprise buildings (SME), CER-IRISH, The smart meter London data-set, Turkish electricity data	Load forecasting based DL model	Stacked Dilated CNN	-	[83]
Australian Government project, Smart Grid Smart City (SGSC)	Power load forecasting of similar-profile energy customers based on Clustering	Pyramid CNN, DBSCAN algorithm, K-means	Short-term	[84]
Power load of a region in New England and a region in Zhejiang, China	Multi-model fusion for load forecasting based on Feature Selection	RF, CNN, BiGRU	Short-term	[85]
Two laboratory buildings and an office building in Switzerland	Building load forecasting model	CNN based on ResNet	Short-term	[86]
American Electric Reliability Council of Texas (ERCOT) power grid	A novel DL and data processing ensemble model (SELNet) for load forecasting	VMD, 2D-CNN, TCN	Short-term	[87]

Case study	Contribution	Approach	Horizon	Ref
The energy consumption data-set for Trento, housel-gold and office data-set	Load Forecasting Based on Data Transformation and Statistical ML	CNN, LSTM	Short-term	[88]
The operation data of the transformer located in Fuzhou	Load Forecasting for transformers in distribution grid based on DL	CNN, LSTM	Short-term	[89]
The Romanian power system	Model implementation for load forecasting	CNN	Short-term	[90]
AMPds2 smart meter data-set	Residential load forecasting based on smart meter data	TCN	Short-term	[91]
Irish Smart Metering Electricity Customer Behaviour Trials (CBTs)	Residential electricity consumption forecasting	CNN	-	[92]
Independent System Operators in New England	A load range discretization method to generate load probability distributions	CNN	-	[93]

Case study	Contribution	Approach	Horizon	Ref
Energy generation and hourly electricity load of the ISO New England Control Area	Load forecasting for residential buildings in smart grids	CNN	Short-term	[94]

Table 4.1: Summary of the state of the art papers presented using CNNs.

4.3 Load Forecasting Based on Recurrent Neural Networks

In the publication by [95], they introduce the use of Bi-LSTM for STLF within a rural microgrid located in Sub-Saharan Africa. Training of the Bi-LSTM involves input variables aimed at predicting the hourly load of the micro-grid by identifying consumption patterns. The outcomes validate the effectiveness of this approach, demonstrating a strong correlation coefficient in comparison to alternative methods employed for STLF.

The authors of this study [96] introduce a novel hybrid model utilizing parallel LSTM-CNN architecture for STLF. The effectiveness of the proposed model is evaluated using the two-hourly load consumption data from Malaysia and the daily power consumption data from Germany. The model leverages the previous consumption as a parameter to forecast the load for the subsequent time step. Its architecture comprises two paths: the CNN path for extracting input data features and the LSTM path for capturing long-term dependencies within the input data. Subsequently, the outputs of these paths are merged, and a fully connected path, along with an LSTM layer, is employed to process the output and predict the final outcome.

In their publication [97], a new model called Singular Spectrum Analysis LSTM was introduced. The forecasting process involves two stages: firstly, SSA is utilized to break down, categorize, and reconstruct the initial load time series. Secondly, the final load is forecasted by LSTM based on the

resulting series. The effectiveness of this approach has been assessed using the Kaggle load dataset and five AEMO electricity load demand datasets.

The research in [98] utilizes models based on DL, ML, and optimization techniques for STLF and MTLF. A three-step model is proposed, comprising feature selection, extraction, and classification. To select the most relevant and important features, a hybrid of Random Forest and Extreme Gradient Boosting is employed to calculate features' importance. The Recursive Feature Elimination method is used to eliminate irrelevant features during the feature extraction process. Load forecasting is carried out using Support Vector Machine SVM and a hybrid ofGRU and CNN. Additionally, metaheuristic algorithms, Grey Wolf Optimization and Earth Worm Optimization, are used to fine-tune the hyper-parameters of SVM and CNN-GRU, respectively.

In their paper [99], the authors introduced a neural network architecture with attention mechanism, GRU, and Bayesian optimization for precise STLF. The attention layer prioritizes essential input data features to enhance accuracy in load forecasting. Bayesian optimization is utilized to fine-tune the model's hyperparameters for optimal predictions. Experimental validation using real power load data from American Electric Power confirms the model's effectiveness in accurate 24-hour load forecasting and algorithm stability.

In [100], a novel ECP framework with three tiers is introduced. The initial tier involves providing the input data sequence to a pre-processing layer to eliminate noise and outliers. The refined data is then fed into the training phase for learning in the second tier, while the third tier generates the final ECP by visualizing the data and results through actual and prediction graphs. This allows for the analysis of data patterns to obtain better interpretations.

In the study conducted by [101], an application utilizing DL methods was created for predicting electricity demand. The authors explored the use of LSTM, CNN and MLP models. They used a typical university electricity consumption dataset to illustrate the impact of weather on electricity loads, particularly in tropical regions. The study compared the performance of these three techniques and concluded that the LSTM model outperformed the others.

A novel approach for forecasting electricity prices and consumption is suggested, utilizing a combination of WT, FS technique, and LSTM network. The WT is used to break down the load and price fluctuations into sub-

series, while the FS module identifies effective inputs for STLF and STPF. Subsequently, the outputs from these steps are fed into the LSTM module for further analysis. The effectiveness of this method was assessed using actual load and price data from the PJM electricity market spanning from 2006 to 2018, price data from the Spanish electricity market in 2002, and load data from Hormozgan province in 2009 [102].

The study [103] involved estimating the hourly load demand using historical 24-hour consumption and weather data (temperature, humidity, wind speed, and radiation) in Konya from 2016 to 2020 to predict the next hour's consumption. Forecasting models utilizing DL algorithms such as RNN, LSTM, and GRU were developed. Among the three models, RNN was identified as the most accurate based on MSE, RMSE, and MAE comparisons.

The authors utilized the k-means clustering method in [104] due to its minimal error rate. They grouped the base stations into five clusters for separate training and testing. The dataset was subjected to analysis using ML and RNN. The performance of LSTM and GRU outperformed statistical methods and ML algorithms in terms of RMSE. The results indicate that using clustered base stations is more effective than using all stations. To compare the tested algorithms, Amazon DeepAR was employed as a benchmark. The results revealed similar error values for both groups. The study utilized the city cellular traffic dataset, a publicly available dataset that presents traffic load statistics for 13k base stations in China.

The authors presented Online Adaptive RNN in [105], which is a method for load forecasting that involves continuous updating of the model as new data becomes available. This approach utilizes an RNN as the base learner to capture temporal dependencies, and it incorporates pre-processing, buffering, and tuning modules for continuous learning. The pre-processing module prepares the data for online learning, the tuning module adjusts the ANN hyper-parameters to accommodate new patterns, and the buffering module assists in learning from challenging patterns and handling concept drift. Online Adaptive RNN outperformed five other online algorithms in terms of accuracy across all forecasting horizons.

The paper cited as [106] predicts the electricity demand for the next five years of G-20 Members by employing GRU, LSTM, Bi-LSTM, and Convolutional LSTM. With the exception of the best model, the error rates of all the models are relatively close. However, based on the overall test error rates, it can be concluded that using a window size of 6 yields the best results when utilizing the LSTM based model.

In [107], a multi-energy load prediction model for RIES was developed using a CGRU hybrid network, HUMTL, and an ensemble approach based on GBRT. The CGRU hybrid network was constructed to extract high-dimensional temporal and spatial features and dynamically model nonlinear time series. By designing three GRUs with different structures, the model can effectively learn various energy features to meet the prediction requirements for different types of loads. HUMTL optimizes prediction tasks for different load types by utilizing homoscedastic uncertainty. Additionally, the ensemble approach based on GBRT allows for sharing prediction results of different structured networks, and the shared results are weighted to obtain more accurate predictions for multi-energy loads. The proposed HUMTL-CGRUG model approximates the evolution laws of different load types, explores the temporal and spatial correlation of multi-energy loads, and delves deeper into the coupling relations among various energy systems compared to other prediction models. This model offers the highest prediction accuracy and the best applicability.

The investigation in [108] focused on examining various models such as Stacked Bi-GRU, Convolutional LSTM, stacked Bi-LSTM Auto-encoder, Bi-LSTM Auto-encoder, hybrid CNN-LSTM, and LSTM-Auto-encoder for hour ahead load forecasting. Based on the findings, the Bi-LSTM-Auto-encoder and CNN-LSTM models were identified as preferable due to their accuracy and training time, respectively. The comparative analysis demonstrated that the CNN architecture significantly reduced the training time by simplifying the model's complexity while improving accuracy compared to the Bi-LSTM model. The best results were achieved through the auto-encoder approach.

The load forecasting model proposed in this research work [109] is based on a Bi-GRU. It incorporates meteorological factors and the impact of holidays to enhance the precision of forecast outcomes. By using the power load data from a district in a Chinese city as an illustration, the model's prediction results fulfill the actual requirements and demonstrate superior prediction performance compared to Bi-LSTM, LSTM, GRU, and other models.

The authors in [110] discuss various methods for predicting power consumption a week in advance and explore the factors that could affect the outcome. Their experiments reveal that the forecast timeframe has an impact on the model's accuracy, with the period from Thursday to Wednesday producing the best results. Furthermore, substituting the day of the week for national holidays is a critical factor. Among several models, the LSTM demonstrated superior performance. It's important to note that due to Tai-

wan's climate, predicting beyond May becomes more challenging.

In this work [111], a power load forecasting model was introduced, which is grounded in the cyberphysical-social systems framework and integrates regional environmental protection policies into the complete load forecasting system. The model was trained using LSTM with the daily load data of 4 provinces in Central China spanning from 2015 to 2018. The findings indicate that the electricity load forecasting model, based on the suggested framework, demonstrates improved accuracy and wider applicability.

The LSTM model, as mentioned in reference [112], forecasted the building and campus load for 18 hours in advance. The input variables taken from the SRRL dataset include Relative Humidity, Barometric Pressure, Dry Bulb Temperature, Global Horizontal Irradiance, Total Cloud Cover, and Wind Speed.

[113] presented a technique based on LSTM for STLF and anomaly correction. They utilized the dataset of hourly power consumption in Toronto, Canada from January to July 2016. This method can adaptively rectify anomalies when faced with missing or min-max data. The difference between the forecasted result over the corrected data and the standard data is minimal. The results suggest that this approach effectively captures the correlation of load series and generates accurate predictions within a short convergence period.

The load forecasting method proposed by [114] is based on LSTM and utilizes the hourly historical system load data of the Electric Reliability Council of Texas (ERCOT) control area spanning from 2003 to 2016. In order to capture the long-term dependencies within the sequence data, the method identifies the intrinsic variation of the load from both the horizontal and vertical dimensions over an extended historical time period, while taking into account various influencing factors. The precision of the LSTM model is demonstrated through the use of actual load data to verify the forecasting performance of different historical date windows and various network architectures in the experimental results.

The authors referenced in [115] utilized historical distribution transformer load data to employ LSTM for predicting future loads. They implemented integrated learning to enhance the model's accuracy and enriched the input dimension to address the challenge of insufficient historical data for model construction. According to the experimental findings, the forecasting approach consistently outperformed alternative methods, offering potential to monitor distribution transformer operation and uphold the health of the dis-

tribution network.

The authors in [116] propose a hybrid model that combines DL pooling LSTM-CNN to forecast short-term and mid-term loads, STLF and MTLF, while also incorporating weather information and time delays. Once the data has been cleaned and pre-processed, the CNN module is employed to capture the sequence of data points, along with different model configurations preceding the LSTM.

In this study [117], the authors introduced a multi-task learning approach for STLF that utilizes Cluster-based Aggregate Forecasting with a diverse dataset of raw load profiles. The suggested method demonstrates good scalability as the number of clusters increases, and the empirical findings significantly support the implementation of this approach.

In [118], three methods were introduced to predict future grid conditions using a DANN architecture based on a learned representation of numerical weather predictions, specifically employing an LSTM model and an Autoencoder. The findings suggest that this approach plays a crucial role in advancing the smart power grid.

This research [119] introduces a forecasting technique that utilizes DL combined with Micro-clustering. This approach is designed around hybrid clustering tasks that are unsupervised and supervised, employing K-means and Gaussian SVM, respectively. By implementing Micro-clustering on the input sequence, the hourly input data is organized into distinct groups. The Bi-LSTM is suggested as the forecasting model. The ideal number of clusters for each hour is established using the Davies-Bouldin index. The study examines forecasting related to wind speed, load demand, and electricity prices across various periods, using data from the Ontario province. The integration of Micro-clustering and Bi-LSTM networks considerably enhanced the forecasting outcomes, particularly during spike occurrences.

A systematic experimental approach has been carried out to explore how stacked LSTM and Bi-LSTM networks affect the prediction of electricity load consumption. Specifically, two stacked configurations consisting of 2 and 3 LSTM layers are compared against a single-layer LSTM for both types to demonstrate the notable significance of incorporating stacked layers. The findings revealed that the stacked LSTM layers did not lead to a considerable enhancement in prediction accuracy. However, these layers required nearly double the processing time compared to the single-layer models. Additionally, the Bi-LSTM networks surpassed the LSTM networks in terms of RMSE values across the 1, 2, and 3 layer model configurations. Moreover,

when comparing prediction accuracy throughout the assessed period, the optimized Bi-LSTM model exceeded both the optimized LSTM model and the SVR model [120].

Table 4.2, lists the papers that tackle the load forecasting problem using RNNs.

Case study	Contribution	Approach	Horizon	Ref
A microgrid in rural Sub-Saharan Africa	DL for load forecasting in a rural microgrid	Bi-LSTM	Short-term	[95]
Hourly load consumption of Malaysia, daily power electric consumption of Germany	Load Forecasting using new hybrid DL model	CNN, LSTM	Short-term	[96]
Australian Energy Market Operator (AEMO) repository data-sets i.e., New South Wales (NSW), South Australia (SA), Tasmania (TAS), Queensland (QLD), Victoria (VIC)	Power load forecasting based on DL models	SSA, LSTM	Short-term	[97]
The latest electricity daily load data-set downloaded from the ISONE website	Big Data analytics forelectricity load forecasting Using an AI techniques ensembler	SVM, GRU, CNN	Short-term, mid-term	[98]

Case study	Contribution	Approach	Horizon	Ref
Real power load data from American Electric Power cpmpany (AEP)	DL forecasting method for electric power load	GRU	Short-term	[99]
Household power consumption data-set publicly available on UCI ML repository	Comparative analysis of the conventional and sequential learning algorithms for electricity load forecasting	ELAs, LSTM, Bi-LSTM, and M-LSTM	Short-term	[100]
Transmission Company of Nigeria	A DL model for electricity demand forecasting	LSTM, CNN, and MLP	Short-term	[101]
Load and price data collected from the PJM electricity market, the price data of the Spanish electricity market, the load data of Hormozgan province in Iran and the Spanish electricity market	Electricity load and price forecasting	LSTM	Short-term	[102]
Hourly data of Konya	Multivariate load forecasting using DL algorithms	RNN, LSTM and GRU	Short-term	[103]

Case study	Contribution	Approach	Horizon	Ref
The city cellular traffic data-set (china)	Forecasting the network traffic load prediction on base stations	LSTM, GRU	-	[104]
The real world data from 5 residential consumers provided by London Hydro	DL for load forecasting	RNN	Short-term	[105]
Yearly electricity domestic consumption data collected by Enerdata organization of all G-20 members	Electricity load forecasting of G-20 members	GRU, LSTM, Bi-LSTM, ConvLSTM	Long-term	[106]
The main campus of the University of Texas at Austin	Multi-energy load prediction model for regional integrated energy systems	CNN, GRU	-	[107]

Case study	Contribution	Approach	Horizon	Ref
ISO New England control area and its eight wholesale load (ISO-NE) records	Hourly load forecasting	Stacked BiGRU, ConvLSTM, Stacked Bi-LSTM-AE, hybrid CNN-LSTM-AE, and LSTM-AE	Short-term	[108]
Power load data of a district of a city in China	Power Load Forecasting	Bi-GRU	Short-term	[109]
Daily industrial power consumption of Taiwan	Electricity load forecasting modeling	LSTM	Short-term	[110]
The daily load data of four provinces in Central China	Power load forecasting model based on the framework of the cyberphysical-social systems	LSTM	-	[111]
SRRL data-set	DL based load forecasting	LSTM	Short-term	[112]

Case study	Contribution	Approach	Horizon	Ref
Hourly power consumption in Toronto Canada	Electrical load forecasting and Anomaly Correction	LSTM	Short-term	[113]
Hourly historical system load data of Electric Reliability Council of Texas (ERCOT) control area	Power system load forecasting method	LSTM	-	[114]
Load power of 15 distribution transformers in a certain area	Load forecasting method of distribution transformer	LSTM	-	[115]
Electric power consumption data	A novel DL approach for electrical load forecasting	Pooling LSTM-CNN	Short-term, mid-term	[116]
SM data from the town of Arbon in northeastern Switzerland	Multitask learning literature approach for electrical load forecasting	RNN	Short-term	[117]
Regional power grid in central Germany	Forecasting power grid states for regional energy markets	LSTM	Short-term	[118]

Case study	Contribution	Approach	Horizon	Ref
Ontario province, Canada data-set	Three forecasting tasks including the wind speed, load demand, and electricity price	Bi-LSTM	-	[119]
Total load for the control area of Switzerland	Electricity load forecasting	LSTM, Bi-LSTM	Short-term	[120]

Table 4.2: Summary of the state of the art papers presented using RNNs.

4.4 Load Forecasting Based on Deep Belief Networks

In [121], the authors presented a method for short-term energy forecasting for electricity, heat, and gas, utilizing deep multitask learning structured around a deep belief network DBN and a multitask regression layer. The DBN is designed to extract features using an unsupervised approach, while the multitask regression layer facilitates supervised predictions. The comprehensive energy forecasting model encompasses various components, including pre-processing, normalization, input characteristics, training phase, and evaluation metrics, all tailored to practical demand and model integrity. Ultimately, the effectiveness of the algorithm and the precision of the energy forecasts for an integrated energy system in an industrial park were confirmed through simulations based on real operational data from the load system. The encouraging results suggest that deep multitask learning holds significant potential for load forecasting.

This study [122] introduces a multi-energy forecasting framework using deep learning techniques to concurrently estimate the electrical, thermal, and

gas net loads of integrated local energy systems. Initially, the distinct multi-energy demand and generation characteristics of diverse prosumers are qualitatively assessed, and these prosumers are categorized into different groups to enhance the forecasting process through a hierarchical clustering approach. Subsequently, a DBN methodology is utilized to uncover the latent features from multi-energy time series, enabling accurate net-load predictions for various prosumers. Finally, the proposed approach is validated with real data from household-scale prosumers. The comparative analysis reveals the advantages and high forecasting accuracy of the proposed approach and confirms its effectiveness in addressing the multi-prosumer prediction challenge involving multiple energy carriers.

In this study [123], a hybrid STLF approach for electric, thermal, and gas systems utilizing DL is introduced. The DBN serves as an unsupervised learning technique, capable of extracting abstract high-level features, while the multitask regression layer facilitates supervised predictions. Following this, a two-stage load forecasting framework comprising offline training and online prediction is developed. The efficacy of this approach is demonstrated using real data from an integrated energy system, revealing that the proposed DL algorithm exhibits outstanding results in terms of both computational efficiency and prediction accuracy.

In [124], the authors introduced a Bisecting K-Means algorithm for clustering load data, which is subsequently decomposed into multiple IMFs using Ensemble Empirical Mode Decomposition. Candidate features are then identified by computing the Pearson Correlation Coefficient (PCC). Ultimately, a hybrid neural network (HNN) forecasting model is proposed, integrating a DBN with a Bidirectional RNN that includes LSTM and Gated Recurrent Unit GRU components. When compared to the forecasting outcomes of alternative methods, this approach demonstrates a significant enhancement in load forecasting accuracy.

In this research [125], an Extreme Learning Machine based on DBN is utilized for power load forecasting. This approach incorporates ELM as the regression component and harnesses the feature extraction capabilities of DBN alongside ELM's robust generalization performance to enhance the precision of power load predictions. The results indicate that, in comparison to traditional DBN models, this new model more effectively captures the fluctuation patterns of power load data time series, leading to improved prediction accuracy.

The paper introduces STLF [127]. Initially, the electric vehicle user be-

haviors are modeled using the Monte Carlo method. Next, instead of using traditional single forecasting models, the LSTM-DBN model is applied to enhance both forecasting accuracy and computational speed for complex scenarios. Subsequently, a dynamic weight distribution method is introduced to merge the two forecasting outcomes. Ultimately, numerical examples demonstrate that the proposed approach significantly outperforms SOTA methods in load forecasting, with the combined model achieving greater accuracy than conventional methods.

This paper [128] introduces a new Markovian switching consensus-based Distributed DBN model aimed at addressing the STLF problem. The model divides the load dataset into multiple local computing agents, which greatly decreases training time and helps prevent overfitting. Additionally, Markovian switching technology is utilized to enhance the model's stability and robustness. The accuracy of the proposed method is validated using GEFCom 2017 competition and ISO New England load datasets. Experimental results involving four local agents show that the suggested DBN algorithm achieves approximately 19% improved forecasting accuracy compared to a centralized DBN algorithm.

In this research [129], a comprehensive approach for energy system load forecasting utilizing DBN and multi-tasking is introduced. The DBN model, which is grounded in RBM, is employed to extract features from data samples via RBMs located at the bottom tier, while the top layer applies supervised BP neural network regression to generate prediction outcomes. The results of the actual study indicate that the proposed method in this paper achieves a high level of prediction accuracy.

Table 4.3, lists the papers that tackle the load forecasting problem using DBNs.

Case study	Contribution	Approach	Horizon	Ref
historical data-sets of heat load, gas load, and electrical load collected from the industrial-park IES demonstration project of Goldwind Technology Co., Ltd., in Daxing District, Beijing, China	Electricity, heat, and gas load forecasting based on Deep Multitask Learning	DBN	Short-term	[121]
A typical prosumer case in Northern China	DL based multi-energy forecasting framework	DBN	-	[122]
An industrial zone in Guangzhou, which consists of a power system, a thermal system, and a natural gas system	Load forecasting of integrated energy systems based on DL	DBN	Short-term	[123]
Grid data-set (real-time power load data recorded by a power company in Chongqing), public data-set power load forecasting competition organized by the European Network on Intelligent Technologies (EUNITE)	Forecasting method based on EEMD and hybrid neural network	DBN, Bi-RNN	Short-term	[124]

Case study	Contribution	Approach	Horizon	Ref
The PJM posted on its website in 2017 power load data	Research on load forecasting method of large Power Grid	DBN, ELM	-	[125]
The electric load data of Liaoning Power Grid	Load forecasting algorithm based on DL considering the flexibility of EV	LSTM, DBN	Short-term	[127]
GEFCom 2017 competition and ISO New England load data-sets	A novel Markovian switching consensus based Distributed DBN model	DBN	Short-term	[128]
Energy data of a certain place in China	Energy system load prediction method based on DBN and multi-tasking learning	DBN	-	[129]

Table 4.3: Summary of the State of the Art Papers Presented Using DBNs.

4.5 Load Forecasting Based on Auto-encoders, Stacked Auto-encoders and Stacked Denoising Auto-encoders

This study [130] introduces an innovative model leveraging deep AE with localized stochastic sensitivity for STLF. It is capable of learning valuable hidden representations by implementing a perturbation strategy in the Q-neighborhood surrounding the training samples. Consequently, it shows a heightened sensitivity to similar unseen samples and effectively extracts features from historical load data. A nonlinear feed-forward neural network, serving as a regression model, utilizes the hidden layer representations from the last layer for load forecasting. To assess the efficacy of the proposed model, four publicly available electricity datasets from ENTSO-E are utilized.

In this study [131], a novel short-term load forecasting (STLF) model is introduced, which integrates Auto-encoder with Random Forest. The Auto-encoder technique is applied to traditional features, excluding previous electricity usage, to generate a condensed set of features. A forecasting model for electric load is then built using the Random Forest method, incorporating these features alongside past electricity usage. To demonstrate the model's effectiveness, three regression models, as well as additional forecasting approaches, are evaluated using actual electric load data from three different building clusters. The findings indicate that the combination of Auto-encoder and Random Forest achieves superior accuracy when compared to other regression models and existing electric load forecasting methodologies.

In [132], the authors utilized DANN to forecast the nationwide electricity demand in Australia by incorporating socio-economic and environmental variables, with predictions ranging from one month to two years into the future, deploying SAE alongside MLP or cascade-forward MLP. Subsequently, both optimized architectures, with their most effective structures and training methods, were employed using SAE as fundamental components to build DANN. This study demonstrated how SAE can enhance the performance of traditional neural network models by uncovering valuable latent features from the original data.

In [133], a study on electrical STLF is conducted through the combination of SAE and GRU. Initially, the SAE extracts features from historical data, which encompasses power load, weather conditions, and holiday information,

followed by the utilization of GRU to build a model for predicting power load. The experimental results indicate that the SAE-GRU model significantly outperforms traditional SVM and GRU models in terms of accuracy for load forecasting, and its prediction time is also shorter than that of conventional LSTM and GRU models.

A hybrid approach is suggested in [134], merging SAE with Extreme Learning Machine techniques. The output generated by each Auto-encoder serves as the input for an individual Extreme Learning Machine. The results achieved are then combined through linear regression to produce the final output. Historical electrical load data was gathered by the Albert Electric System Operator and made available to market participants.

A load data mining approach based on SAE is recommended in [135]. Initially, a novel framework for the data flow from smart meters is created. On the user end, the SAE is employed to derive features from the load data. Subsequently, centralized classification is implemented at a remote data center through a Softmax classifier, with a fine-tuning strategy integrated to enhance both the extracted features and the accuracy of classification. Case studies conducted in China and Ireland illustrate that the proposed technique increased the classification accuracy for both appliance-level and house-level datasets.

In [136] a DL model is proposed. Instead of directly using the original data, a feature extraction module is designed containing multiple SDAE based DANN to refine more abstract features from historical load data and related temperature parameters, and then the resultant features are used as input to an SVR model. Comparative experiments with SVR and ANN demonstrate its effectiveness and superior performance, evidenced by lower MAPE values.

In this study [137], the SDAE is utilized to carry out the STLF task by considering four factors: historical loads, somatosensory temperature, relative humidity, and daily average loads of the Chinese city Fuyang. The model based on SDAE demonstrates superior performance in comparison to conventional methods as an innovative approach for STLF. By employing the greedy algorithm for both pre-training and fine-tuning, the SDAE addresses the issues of overfitting and vanishing gradients.

This paper [138] introduces a probabilistic SDAE based method for predicting transient stability to tackle the uncertainties associated with wind farms and loads, incorporating every possible operating scenario as input for the SDAE model. The application of Lasso diminishes the need for training

data without compromising prediction accuracy. The evaluation results on the adjusted IEEE 39-bus system indicate that the suggested model achieves a comparable level of accuracy to MC and LHSMC while offering substantially greater computational efficiency.

Table 4.4, lists the papers that tackle the load forecasting problem using AEs, SAEs ans SDAEs.

Case study	**Contribution**	**Approach**	**Horizon**	**Ref**
Four real world public electricity data-sets from ENTSO-E	Novel model based on deep AE with localized stochastic sensitivity for load forecasting	AE	Short-term	[130]
Typical 15 min interval electric load data of a private university in Korea	Load forecasting based on AE and RF	AE	Short-term	[131]
Monthly record of electricity demand reported by the Australian Energy Market Operator (AEMO)	Electricity demand forecasting using Deep ANN	SAEs, MLP, CFMLP	Mid-term, long-term	[132]
Power data of a certain city in a province	Electrical load forecasting method based on DL	SAEs, GRU	Short-term	[133]

Case study	Contribution	Approach	Horizon	Ref
Historical electrical load data-set collected by the Albert Electric System Operator (AESO), the electrical load data of the Oak Ridge National Laboratory	Electrical load prediction based hybrid model	SAEs, ELM	-	[134]
Electric Data Acquire System (EDAS) of State Grid Zhejiang Elec, Power Corp (SGZEPC) in China, the Sustainable Energy Authority of Ireland (SEA)	Electric load data compression and classification	Deep SAEs	-	[135]
	An efficient DL model for electricity load forecasting	SDAEs	Short-term	[136]
Electric load data of a city in southern China	DL for electric load forecasting	SDAEs	Short-term	[137]

Case study	Contribution	Approach	Horizon	Ref
-	Probabilistic SDAE based transient stability prediction method to address the uncertainties of wind farms and loads	SDAEs	-	[138]

Table 4.4: Summary of the state of the art papers presented using AEs, SAEs and SDAEs.

4.6 Deep Artificial Neural Networks used for Load Forecasting: Algerian Case Study

The PREVOS-DZ forecasting tool was created to forecast national and regional MTLF and STLF, utilizing several Multiple Regression and ANN models in the form of MLP [139].

Researchers in [140] introduce new CNN architectures, such as CNN-1D and CNN-2D, for short-term load forecasting (STLF), utilizing historical load data to identify various factors affecting power consumption. They have found that chronological contexts, such as weekends and holiday events, along with external variables like daily temperatures, are closely linked to power consumption. Thus, these factors were incorporated as inputs into the model. The input format for the CNN-1D is a 1D vector encompassing all the variables, while the input for the CNN-2D is a 2D representation in the form of a matrix created from the chosen features. After evaluating both architectures, the CNN-2D was determined to be the more accurate model when compared with the CNN-1D, in addition to other machine learning models developed from historical load data featuring exogenous variables. The performance of the models was assessed for one-step-ahead forecasting

across both 15-minute and 24-hour intervals. The results from the CNN-2D were notably promising for one-step-ahead forecasts.

An illustration of the application of the SDAE model is found in [141], which utilizes data reflecting the Algerian hourly electricity demand over a four-year span from 2013 to 2016. For each model's inputs, certain load values are selected based on auto-correlation analysis along with one non-linearly associated external variable. Consequently, the electricity load is described by multiple previous values in addition to temperature forecasts. Additionally, the day type context is captured using one-hot encoding for each day of the week, along with an extra input for special occasions, resulting in a total of 16 input variables.

Table 4.5 lists the papers that tackle the algerian load forecasting problem using deep ANNs.

Case study	**Contribution**	**Approach**	**Horizon**	**Ref**
Electricity load in Algeria using the data provided by the SONELGAZ	Algerian Electric Load Forecasting Software	MR, ANN	Short-term, Mid-term	[139]
Electricity load in Algeria using the data provided by the SONELGAZ	A novel DL method for load forecasting	CNN	Short-term	[140]
4 years of electricity data provided by SONELGAZ	Two stage approach for load forecasting using temperature estimation	SDAEs	Short-term	[141]

Table 4.5: Summary of the state of the art papers presented: Algerian case study.

4.7 Conclusion

In this chapter, we attempted to provide an extensive review of electrical load forecasting challenge using DANN architectures over several time horizons. The emphasis is put on the most recent papers presented based on CNN, RNN, LSTM, GRU, DBN, AE, SAE, and SDAE. Furthermore, a review of Algerian load forecasting is presented as a case study. Finally, we summarized each review section into an organized table.

Chapter 5

Electrical Load Forecasting Based on Uni-modal Data

5.1 Introduction

A nation's economic growth, and development are heavily reliant on the supply of electrical energy [142]. This energy must maintain a balance between production and consumption to ensure sustainability. Various elements, including economic activities and climatic conditions, can influence short-term energy consumption levels. In Algeria, electricity usage exhibits a seasonal pattern, peaking in the summer and decreasing in the winter. Therefore, accurately predicting short-term electricity demand is crucial for ensuring that the overall output of the electricity grid maintains a balance between production and consumption throughout the day, thereby preventing both power shortages and excess capacity, which can lead to significant losses.

5.2 Time Series Data

Time series refer to a series of measurements of the same variable recorded at different times, typically at consistent intervals. Depending on the situation, the time variable may be represented in seconds, minutes, hours, days, and so forth. One unique characteristic of time series is that each observation at a specific time is influenced by the values that came before it, which enables us to analyze its progression and predict its future patterns. In models of electrical load time series, the relationship between electricity usage

and various factors is assessed before selecting which exogenous variables are relevant; see Figure 5.1. Auto-regressive variables, i.e. The past observations of the same variable, measured using initial variables that occur in the short-term have been taken into consideration. In Figure 5.2, a noticeable co-movement can be observed between electricity consumption and its Auto-regressive variable represented by the electricity consumption of the previous hour after data normalization.

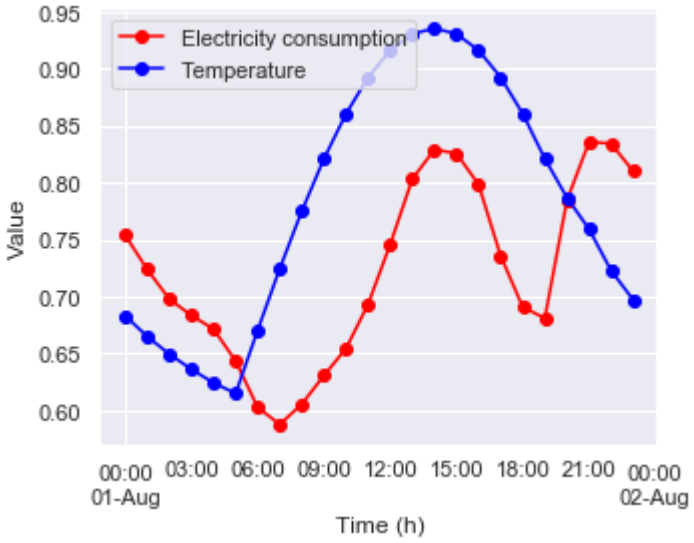


Figure 5.1: Co-movement Between Time Series Data of Electricity Consumption and Temperature.

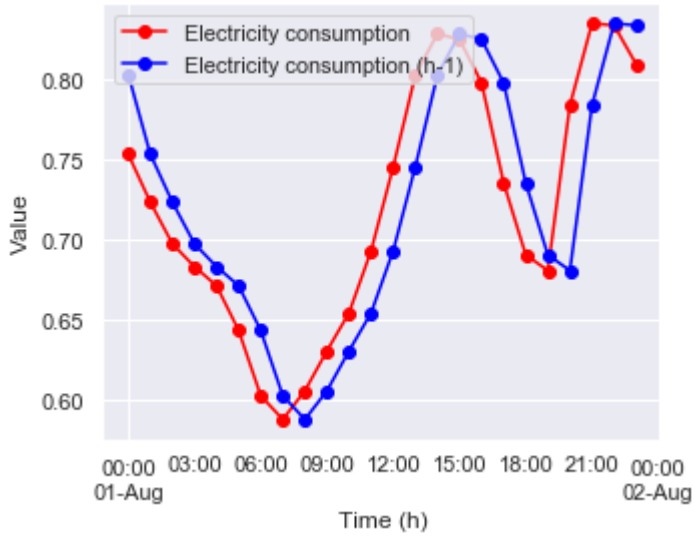


Figure 5.2: Co-movement Between Time Series Data of Electricity Consumption and the Auto-regressive Variable (Electricity Consumption of the Previous Hour).

The maximum and minimum temperature and electricity consumption data is sourced from SONELGAZ. In the original dataset, the input vector includes the temperature factors (both maximum and minimum) along with a one-hot-encoded variable that is 0 on regular days and 1 on special days. The output represents the peak load curve, indicating the highest electricity demand from all consumers connected to the network.

5.2.1 Estimation of hourly temperature profile

Following Linvill's method [146] [141], the daily temperature curves with an hourly resolution are obtained from daily maximum and minimum temperature observations. The temperature wave from sunrise to sunset can be described by Eq. (5.1):

$$T(t) = (T_{max} - T_{min}) \times \sin \frac{(\Pi \times t)}{DL + 4} + T_{min} \tag{5.1}$$

Where $T(t)$ is temperature at time t after sunrise; T_{max} is maximum temperature; T_{min} is the morning minimum temperature, and DL is day length (in hours).

Eq. (5.2) below is used to define nighttime cooling from sunset to sunrise:

$$T(t) = T_s - \frac{(T_s - T_{min})}{\ln(24 - DL)} \times \ln(t) \tag{5.2}$$

where $T(t)$ is temperature at time $t > 1$ hour after sunset and T_s is the sunset temperature obtained from Eq. (5.1). The other terms are as defined in Eq. (5.1). Linvill's equations required some input parameters like sunset time, sunrise time and day length. Those parameters were computed after applying [147] and [148] procedures based on the geographic latitude [141].

In Algeria, the northernmost latitude is around 37° N and the southernmost latitude is around 19° N. So we stopped at 28° N in the middle.

Day length is the time between sunrise and sunset, typically expressed in hours and it depends on:

- Latitude of the location ϕ
- Day of the year (n, where January 1 = 1)

The Day Angle Γ (in radians) is calculated by Eq. (5.3):

$$\Gamma = \frac{2\pi(N-1)}{365} \tag{5.3}$$

The Solar Declination δ (in radians) is calculated by Eq. (5.4):

$$\begin{aligned}\delta = 0.006918 - 0.399912\cos\Gamma + 0.070257\sin\Gamma - 0.006758\cos(2\Gamma) + \\ 0.000907\sin(2\Gamma) - 0.002697\cos(3\Gamma) + 0.00148\sin(3\Gamma)\end{aligned} \tag{5.4}$$

The Sunset Hour Angle ω_s (in radians) is calculated by Eq. (5.5) Eq. (5.6):

$$\cos\omega_s = -\tan(\phi)\cdot\tan(\delta) \tag{5.5}$$

$$\omega_s = \cos^{-1}\left(-\tan(\phi)\cdot\tan(\delta)\right) \tag{5.6}$$

Converting Hour Angle to Time (hours from solar noon), see Eq. (5.7):

$$t_s = \frac{\omega_s \times 180/\pi}{15} \tag{5.7}$$

Calculating Times (Local Solar Time), see Eq. (5.8), Eq. (5.9) and Eq. (5.10):

$$SunriseTime(LST) = 12 - t_s hours \tag{5.8}$$

$$SunsetTime(LST) = 12 + t_s hours \tag{5.9}$$

$$DayLength(hours) = 2 \times t_s \tag{5.10}$$

LST stands for Local Solar Time, it is time measured based on the position of the Sun in the sky at a specific location.

5.2.2 Auto-regressive variables

One of our contributions in this research, is introducing the concept of auto-regressive variables in time series modeling. A variable in time series models can be explained by its Auto-regressive variables, which means that it can be explained by its own past values and by the past values of other variables. The relevant auto-regressive variables that were included in the input vector are:

- Temperature of the previous 24 hours.
- Temperature of the previous 24 hours in the previous week.
- Electricity demand of the previous 24 hours.
- Electricity demand of the previous 24 hours in the previous week.
- Average temperature of each season.
- Season's average temperature of the previous 24 hours.

5.2.3 Data normalization

Data normalization is carried out to adjust the data within a designated range. This procedure guarantees the quality of the data before it is input into any learning algorithm, thereby enhancing the accuracy of predictions [149]. Consequently, a value of X is transformed into X' by utilizing Eq. (5.11):

$$X' = \frac{x}{x_{max}} \tag{5.11}$$

where X' is the new value, X is the old value and X_{max} is the largest value selected based on previous works on mid-term prediction [144].

The shape of the normalized data-set is (28488,125), including the temperature data, electric load, special day variable and Auto-regressive variables of both electric load and temperature variables, see Table 5.1.

It is important to highlight that the information provided in this section is the only available set of data. It is collected during 4 years from 2013 to 2016 and divided sequentially into 60%, 20% and 20% for learning, validating and testing respectively.

	Data	**Description**
Input	*Temperature*	Maximum Temperature Minimum Temperature Hourly temperature Auto-regressive variables: $[T_{h-1}...T_{h-24}]$ Auto-regressive variables: $[T_{h-168}...T_{h-192}]$ The average temperature of the season Ts Auto-regressive variables: $[Ts_{h-1}...Ts_{h-24}]$
	Special day	One hot encoded $\begin{cases} 1 & \text{if special day} \\ 0 & \text{otherwise} \end{cases}$
	Electric load	Auto-regressive variables: $[L_{h-1}...L_{h-24}]$ Auto-regressive variables: $[L_{h-168}...L_{h-192}]$
Output	*Electric load*	Electricity demand

Table 5.1: Summary of Input and Output $_{D}ata$.

5.3 Model Development

5.3.1 Multi-layer Perceptron

The comprehensive typology of the MLP model is presented in Table 5.2 below. The number of units in the input layer matches the size of the input vector, and given that we are engaged in a regression task, the output layer includes only a single unit. The model has two fixed hidden layers, and the number of units in each layer is determined through empirical testing. The non-linearity applied is PReLU Eq. (5.12), while the output layer employs a linear function Eq. (5.13).

$$f(x) = \alpha x \text{ when } x > 0 \tag{5.12}$$

Where α is a parameter, determined by network itself during training.

$$f(x) = ax + b \tag{5.13}$$

Hyperparameter	**Value**
Number of units / layers	125-250-125-1
Activation functions	Hidden layers: PReLU Output layer: Linear
Learning rate	0.001
Loss function	Mean Squared Error
Number of epochs	900
Batch size	128

Table 5.2: Conceptual Parameters for the Multi-layer Perceptron Model.

5.3.2 Stacked Denoising Auto-encoder

The SDAE consists of three DAE, with each encoding layer having dimensions of 50, 20, and 8 respectively. Prior to conducting the unsupervised training, a noise level of 60% is added to the input data. The model's architecture and hyperparameters can be found in Table 5.3.

Hyperparameter	**Value**	
	Pre-training	*Fine tuning*
Number of units / layer	Encoding dimensions: 50-20-8	125-50-20-8-1
Activation functions	Encoding: PReLU Decoding: PReLU Output : Linear	Hidden layers: PReLU Output layer : Linear
Learning rate	0.001	0.001
Loss function	Mean Absolute Error	Mean Absolute Error
Number of epochs	500-500-500	900
Batch size	32-32-32	32
Noise	60%	-

Table 5.3: Conceptual Parameters for the Stacked Denoising Auto-encoder Model.

5.4 Results and Discussion

To evaluate the robustness of the models, MAPE Eq. (5.14) has been employed to evaluate the various architectures.

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{y - \hat{y}}{y} \right| \tag{5.14}$$

Where y is the real output, and $\hat{y}$ is the predicted value.

The experimental findings, based on MAPE values, indicate that the SDAE model slightly outperformed the benchmark scheme MLP, which required Auto-regressive variables, with a MAPE value of 0.95%. In contrast, the MAPE value obtained from the MAPE model was marginally higher at 1.03%, resulting in a difference of 0.08%. It's important to note that both errors fall within the tolerance and specification range set by the national electricity company. However, when we assess the performance of the two models based on training time ratios, it can be concluded that the MLP excelled in terms of training speed.

The comparison between the actual consumption values and the estimated consumption values for three days is shown in Figure 5.3 below. From analyzing the curve, it can be observed that the overall outcomes are quite comparable; however, the SDAE shows superior performance in terms of peak loads. It's important to highlight that the evening peak is the key point for forecasting, as it represents the highest production. The results from the two architectures are summarized in Table 5.4. Subsequently, the model's output is denormalized to align with the actual load values.

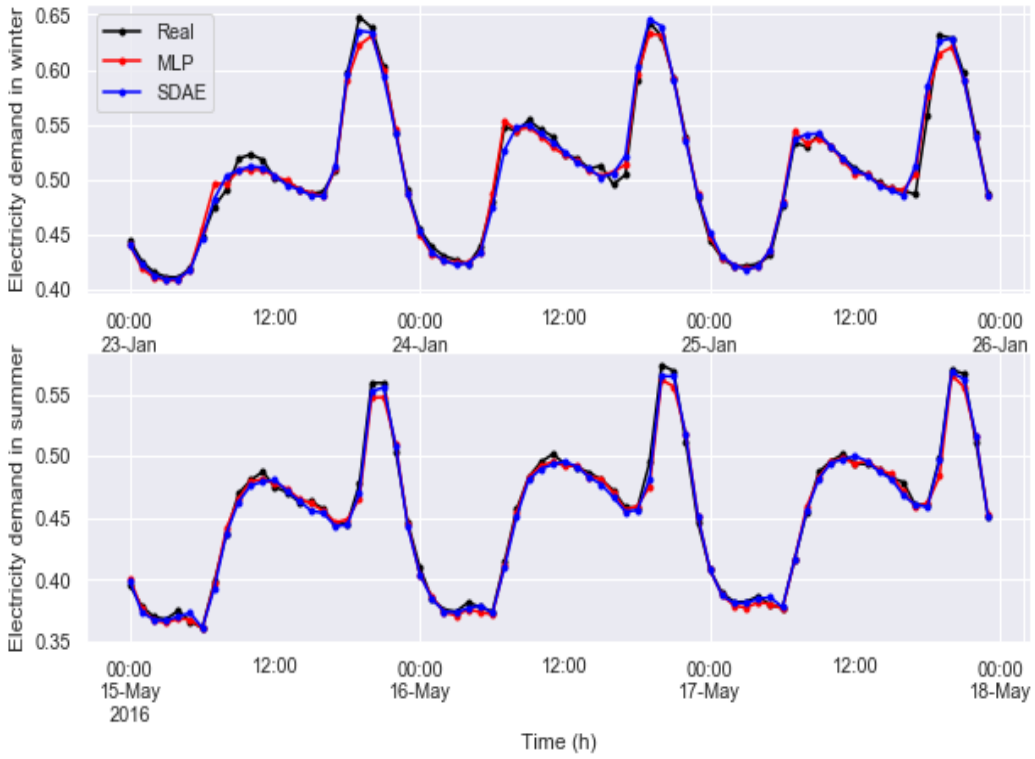


Figure 5.3: Visualization of Multi-layer Perceptron and Stacked Denoising Autoencoders Results Within three days Compared with Real Values.

	Initial Data		**Initial data + auto-regressive variables**	
	MLP	*SDAE*	*MLP*	*SDAE*
MAPE (%)	11.65	11.55	1.03	0.95
Training Time (s)	-	-	411	1918

Table 5.4: Summary of the Results Obtained by the Two Models.

5.5 Conclusion

In this chapter, the forecasting of Algerian electrical load was examined using SDAE. This method was evaluated in terms of MAPE and training duration in comparison to a conventional benchmark MLP model. Indeed, the SDAE did not significantly outperform the traditional MLP, with the exception of slightly better alignment with evening peaks, which led to a modest improvement in MAPE. This improvement came at the cost of increased computational time (by a factor of 5). SDAE are worth exploring for electrical load forecasting when the data is expanded to include multi-modal one.

Chapter 6

Analysis of the Impact of Multi-modality on Sort-term Electrical Load Forecasting

6.1 Introduction

Grid technology plays a crucial role in economic growth and development. Its expansion continues substantially, even as individuals seek alternative energy sources. Consequently, load forecasting serves as a vital instrument for formulating effective scheduling strategies employed by both public and private energy suppliers to achieve an optimal balance between supply and demand. This guarantees sustainability regarding energy requirements and avoids mistakes that could lead to excess costs [142]. Load forecasting can be divided into three categories:

STLF: Several minutes to one day ahead prediction, it is essential for power suppliers to operate power plants and is the subject of the present study.

MTLF: Used for resource planning.

LTLF: Used by decision-makers from industry and other interested stakeholders to plan and build power supply and transmission facilities [142, 139].

In Algeria, electricity consumption exhibits a seasonal pattern, peaking during the summer months and decreasing in winter, which clarifies the short-term load reaction to various influences like weather conditions, such as temperature, and economic activities, e.g. holidays [143]. Consequently, exam-

ining how electricity demand reacts to these factors is essential for precise demand forecasting.

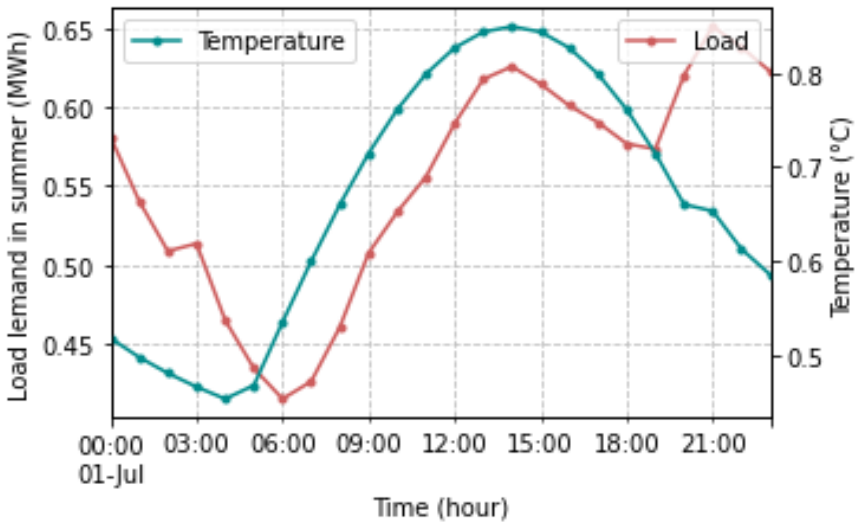
Individual modalities have been until recently, the sole employed approach in load forecasting. In this context, this work's contribution lies in presenting the concept of multi-modality to address the load forecasting challenge. This concept is introduced to highlight the collaborative aspect of various modalities that handle the same phenomenon.

6.2 Methodology and Experiments

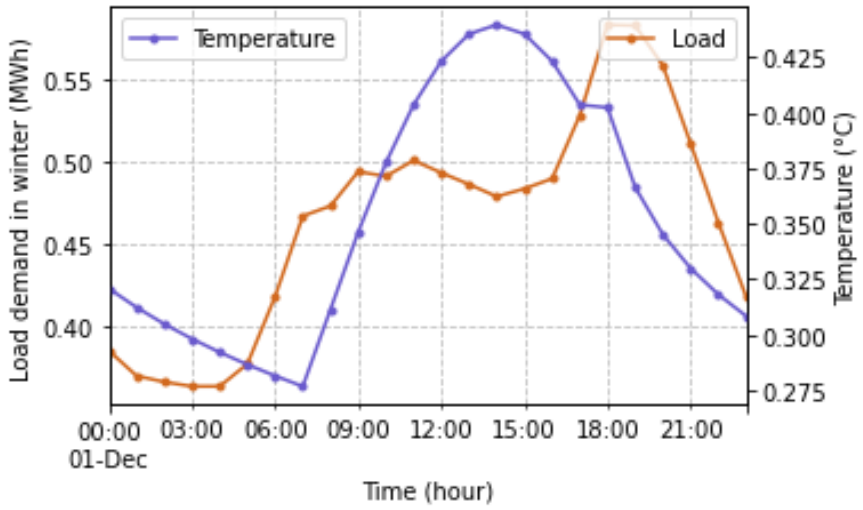
6.2.1 Multivariate time series

Time series data consists of sequential observations of any kind of information. What sets it apart from other data types is the possibility of relationship between observations; the value of each data point depends on its preceding values. This allows for an analysis of past trends to forecast future behavior.

The influence of temperature on STLF is substantial and is regarded as the most significant external factor [17, 18, 19]. The co-movement between electricity usage and temperature offers an initial investigation into the connection between these time series, which can be seen in Figure 6.1, where the electrical load curves for two typical working days in Algeria, one in winter (2014-12-01) and another in summer (2014-07-01), are compared against the temperature curve for the same days.



(a) Load curves of Algeria in (2014-07-01).



(b) Load curves of Algeria in (2014-12-01).

Figure 6.1: Electricity Load Curves of Algeria in a Typical Day in Winter and Another in Summer.

The only exogenous variables initially provided by the System Operator

of SONELGAZ are limited to daily maximum and minimum temperature values. Alongside temperature data, load demand is reported, with special emphasis placed on the daily peak, which indicates the highest electricity demand from all consumers linked to the network. This study utilizes a comprehensive data set that includes recordings over a three-year period, from 2013 to 2016.

Hourly temperature profile estimation

The hourly temperature data points were derived from daily maximum and minimum temperature observations to strengthen our input vector with additional important factor, following Linvill's method [19, 141].

The temperature wave from sunrise to sunset can be described by Eq. (6.1):

$$T(t) = (T_{max} - T_{min}) \times \sin \frac{(\Pi \times t)}{DL + 4} + T_{min} \tag{6.1}$$

where $T(t)$ represents the temperature at time t following sunrise, T_{max} indicates the maximum temperature, T_{min} corresponds to the minimum temperature in the morning, DL stands for the duration of the day in hours and $\Pi = 3.14159$. Eq. (6.2) defines the nighttime cooling beginning at sunset:

$$T(t) = T_s - \frac{(T_s - T_{min})}{\ln(24 - DL)} \times \ln(t) \tag{6.2}$$

Where $T(t)$ represents the temperature at time $t > 1$ hour after sunset, and T_s denotes the sunset temperature obtained from Eq. (6.1). The other variables are defined in Eq. (6.1). Linvill's equations require several input parameters, such as sunset time, sunrise time, and day length. These parameters are calculated following the procedures outlined in [20] and [21] based on the geographic latitude [141].

Auto-regressive variables

Our contribution is to introduce the concept of multivariate time series. In multivariate time series models, an endogenous variable y can be explained by the auto-regressive variables, in another words, by its own past observations as well as the past observations of other exogenous variables $(x_0, x_1, \ldots x_j)$.

The auto-regressive variables incorporated into the input vector are outlined below:

- Temperature profile of the previous 24 hours.
- Temperature profile of the 24 hours of the same day in the previous week.
- Load profile of the previous 24 hours.
- Load profile of the 24 hours of the same day in the previous week.

6.2.2 Data normalization

ANN handle inputs by applying small weights, and inputs with high values can disrupt or hinder the learning process. Consequently, normalizing data is an important step. This involves adjusting the values so that each one falls within the range of 0 to 1 [25] to guarantee a suitable data quality before it is used in any learning algorithm.

Data normalization can be achieved by dividing all data by the largest existing value, see Eq. (6.3). Whereas in the case of image data, the matrices of pixel values are divided by 255 which represents the largest pixel value. This is performed across a single channel as we are processing gray-scale images to reduce the complexity of the model.

$$X' = \frac{X}{X_{max}} \tag{6.3}$$

X is normalized to X', where X' is the new value, X is the old value and X_{max} is the largest value.

The summary of all the modalities is shown in Table 6.1;

Modality	Description
	Gray-scale image representing the load curve of the previous 24 hours in the shape of 32x32.

Temperature and load data	• Maximum temperature of the day, • Minimum temperature of the day, • Hourly temperature, • Average temperature of the season, • Temperature profile of the previous 24 hours, • Temperature profile of the 24 hours of the same day in the, previous week, • Load profile of the previous 24 hours, • Load profile of the 24 hours of the same day in the previous week.
Type of the day	One-hot-encoded vector of shape (11x1).

Table 6.1: Summary of all the Modalities.

6.2.3 Model Development

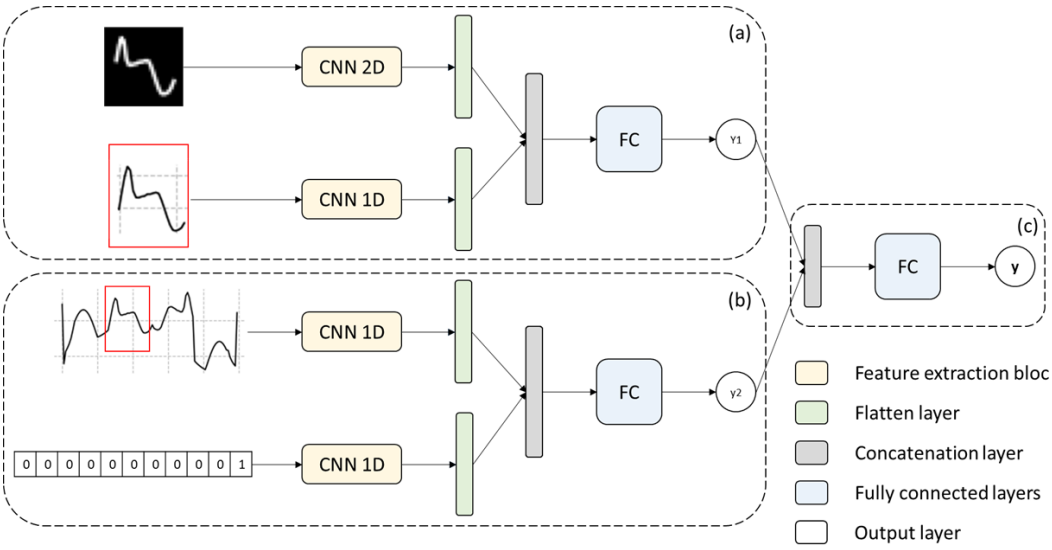


Figure 6.2: Architecture of the Proposed Multi-modal Deep Learning Approach Using a Combination of One Dimensional Convolutional Neural Network and Two Dimensional Convolutional Neural Network Models.

CNN-2D-CNN-1D Framework

Given the presence of diverse data types, the suggested framework depicted in Figure 6.2 adheres to a domain-specific ANN. It conducts separate processing for each modality to produce their feature representations using distinct DANN based on the input data type. The CNN-2D module is responsible for extracting features from the image data, while the CNN-1D module handles the processing of one-dimensional time series data.

In the sub-network (a), two modalities were processed: an image of the load curve from the past 24 hours and a vector representing the load curve from the previous 24 hours.

The first sub-model consists of a CNN-2D. It is created to extract features from the image modality, utilizing three convolution layers that employ 3x3 sized kernels. Each layer contains a varying number of kernels: 16, 8, and 1 in succession. Additionally, a 2x2 max-pooling layer is added at the end of the convolution block to halve the size of the feature map.

The second sub-model used a CNN-1D architecture. The same segment of one-dimensional data that was employed for generating the image modality serves as the input vector for the CNN-1D sub-model. We have implemented three convolution layers with a kernel size of 3, with the number of kernels in each layer set to 32, 16, and 8, respectively. The padding parameter is consistently fixed to "same" across all layers.

At the top of each sub-model, flatten layers were added. The generated outputs were combined and passed to a second-level model to obtain a first-level output. The structure of the fully connected neural network is designed using 10 neurons in the hidden layer and 1 neuron in the output layer.

The vector representing the load curve of the previous 24 hours previously processed in the sub-network (a) is also integrated in the third feature extraction bloc to contribute in the extraction of a meaningful features when combined with another modality.

The third sub-model consists of a CNN-1D architecture, taking the second modality as input which consists of a vector of numerical values shaped (100,1).

The sub-model is CNN-1D consisting of a 1 convolution layer with 16 filters of size 3, followed by an average-pooling layer and finally a flatten layer.

Following the same process of merging the layers in the previous sub-network (a), the output layers of the third and fourth sub-models are combined with a second intermediate fully connected neural network, with 30, 8, and 1 neurons in each respective layer.

In this phase, the results of (a) and (b) are passed to a meta model with two hidden layers consisting of 30 and 8 neurons, respectively, and a single-neuron output layer to make predictions.

6.3 Experimental Results and Discussion

For the purpose of this research, the full dataset collected over a three-year period from 2013 to 2016 and sampled with an hour interval were divided into three segments: 60%, 20%, 20% batches for training, validating and testing, respectively.

The experiments aimed at addressing the STLF task were carried out to showcase the effectiveness of the proposed approach, focusing on both the model's accuracy and the training duration of the various methods. It is

important to highlight that the numerical calculations were conducted on a computer equipped with a CPU, which necessitated a significant amount of time for training the DL architectures, unlike GPU computing. Nevertheless, the CPU operating at 2.50 GHz and a RAM capacity of 4.00 GB proved adequate for carrying out the training and validation processes.

Below are the experimental results derived from three different architectures applied to three unique case studies, based on Eq. (6.4), which depicts the MAPE. These results are intended for comparison and further discussion.

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{y - \hat{y}}{y} \right| \tag{6.4}$$

Where y and $\hat{y}$ refer to the real output and the predicted value respectively.

6.3.1 First case study: Initial data

The initial dataset is limited to temperature-related information, specifically the maximum and minimum temperatures recorded for each day, as well as the estimated hourly temperatures. The results derived from various models are presented in Table 6.2 below.

Model	MAPE (%)	Training time (s)	Prediction time (s)
MLP	8.35	297.61	0.84
SDAEs	9.93	425	0.59
CNN-1D	8.12	440.04	0.76

Table 6.2: Results Based on Initial Data.

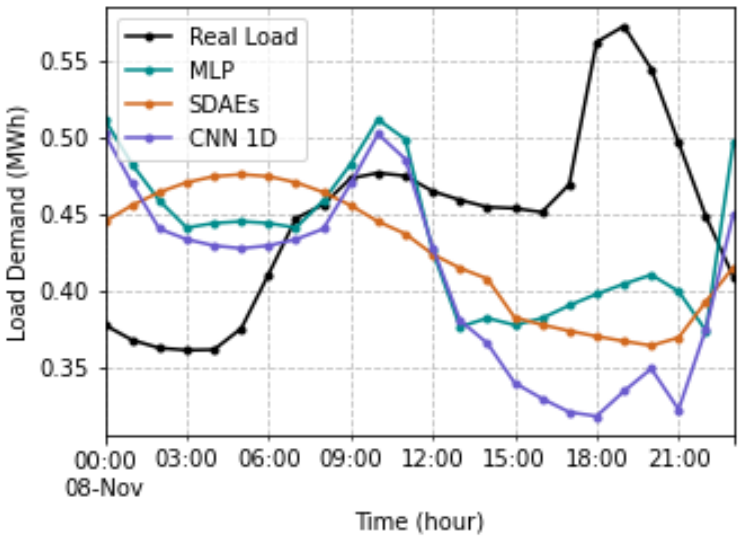


Figure 6.3: Estimated Load Using Initial Data Compared to Real Load Values.

The results obtained by the MLP, SDAE and CNN-1D models based on initial dataset are illustrated in Table 6.2 highlighting a considerable lack of precision regarding STLF when a small input vector is used, represented by temperature factors. The results are visualized in Figure 6.3.

The CNN-1D model achieved a MAPE with a value of 8.12%. In contrast, the MLP model achieved a MAPE of 8.35%. The fact that SDAE architecture is more suitable for large input vectors since it performs a dimensionality reduction over the input vector of each layer, the prediction error of SDAE model is the worse with a value of 9.93%.

When assessing training time, the MLP model is the most efficient, owing it to its simple architecture in comparison with the CNN-1D and SDAE models. However, in terms of prediction time, the SDAE exhibited superior speed compared to the other models. It should be emphasized that the training duration for the SDAE in all examined case studies is limited to the fine-tuning stage.

It can be inferred from the results that the STLF cannot rely only on temperature related data in the case of small input vector.

6.3.2 Second case study: Integration of auto-regressive variables

By integrating the auto-regressive variables, the dimensions of the initial input vector have been extended to (100,1). The consideration of these auto-regressive variables aims to evaluate their effect on the model's accuracy in terms of generalization capacity. The summarized results can be found in Table 6.3 and are depicted in Figure 6.4.

Model	MAPE(%)	Training time (s)	Prediction time (s)
MLP	1.02	654.06	0.41
SDAE	0.99	755.10	0.37
CNN-1D	0.90	1109.64	1.12

Table 6.3: Results obtained by the different models using initial data in addition to auto-regressive variables.

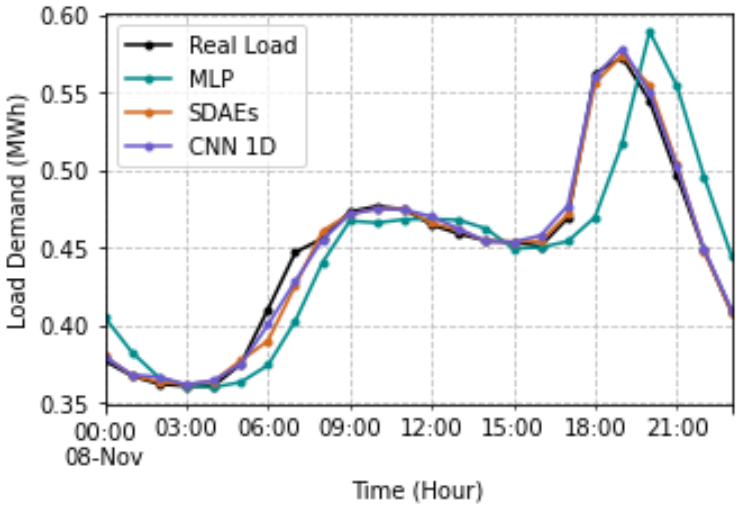


Figure 6.4: Estimated Load Using Auto-regressive Variables Compared to Real Load Values.

The results depicted in Figure 6.4 demonstrate a significant enhancement in the predicted values across all architectures when the auto-regressive variables are incorporated, as opposed to the initial experiment. Furthermore, Table 6.3 highlights the exceptional performance of the CNN-1D model, which achieves the lowest MAPE value of 0.90%. However, it is noteworthy that both the training time and prediction time for this model are greater than those of the other architectures.

As indicated in Table 6.3, the MLP and SDAE models exhibit comparable performance levels. The prediction errors for the MLP and SDAE are 1.02% and 0.99%, respectively. The difference in training and prediction times between these models is negligible, with an observed improvement in prediction time relative to the previous experiment. The overall findings of this case study reveal a notable enhancement in prediction error across the three architectures when the input data is expanded beyond only the temperature factor. Additionally, they underscore the significance of the input vector when integrating auto-regressive variables pertaining to both temperature and load data.

6.3.3 Third case study: Multi-modal data

In this phase, which is the subject of this study, all modalities are integrated to highlight their complementary and collaborative characteristics in the estimation of the STLF. The modalities incorporated into the time series modeling consist of the temperature and load time series vectors, the image curves of the load, and the one-hot-encoded vector that denotes the type of day.

Model	MAPE (%)	Training time (s)	Prediction time (s)
CNN-2D-MLP	0.98	10312.96	61.05
CNN-2D-SDAE	0.96	8911.58	5.63

CNN-2D-CNN-1D	0.87	4033.17	6.48

Table 6.4: Results Obtained by the Combined Models Based on Multi-modal Data.

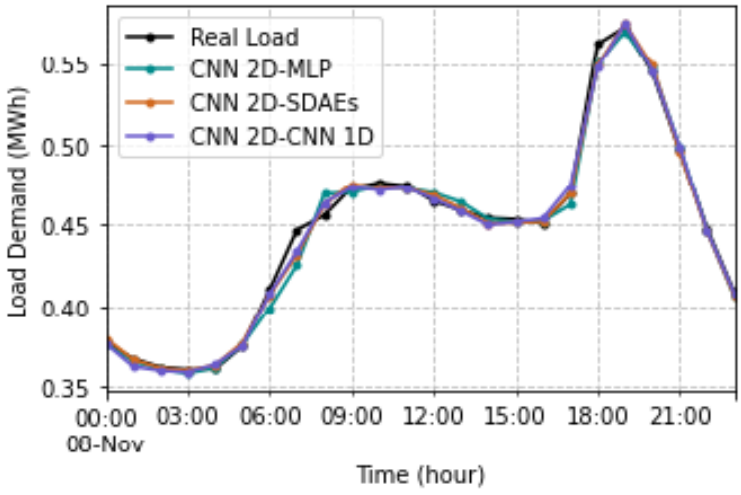


Figure 6.5: Estimated Load Using Multi-modal Data Compared to Real Load Values.

The results presented earlier in Table 6.4 illustrate the effectiveness of the models' fusion method in handling multi-modal data for the purpose of tackling the STLF. Figure 6.5 reveals an increased accuracy in the predicted values when juxtaposed with the actual load values, particularly in relation to the earlier case studies.

The combined architecture CNN-2D-CNN-1D performed the best with a MAPE value of 0.87%, followed by CNN-2D-SDAE and CNN-2D-MLP architectures with MAPE values of 0.96% and 0.98% respectively.

With regard to training time, the CNN-2D-CNN-1D architecture converged faster than the other architectures being the optimal solution with the lowest prediction error. Although there is an increase in training time

in comparison to the previous case studies due to the complexity of the architecture used to process multi-modal data, the models have improved in terms of accuracy.

In terms of training duration, the CNN-2D-CNN-1D architecture is the optimal solution, it demonstrated a quicker convergence compared to the other architectures. While there is a notable increase in training time relative to earlier case studies, attributed to the complexity of the architecture designed for multi-modal data processing. However, the models have shown enhancements in accuracy.

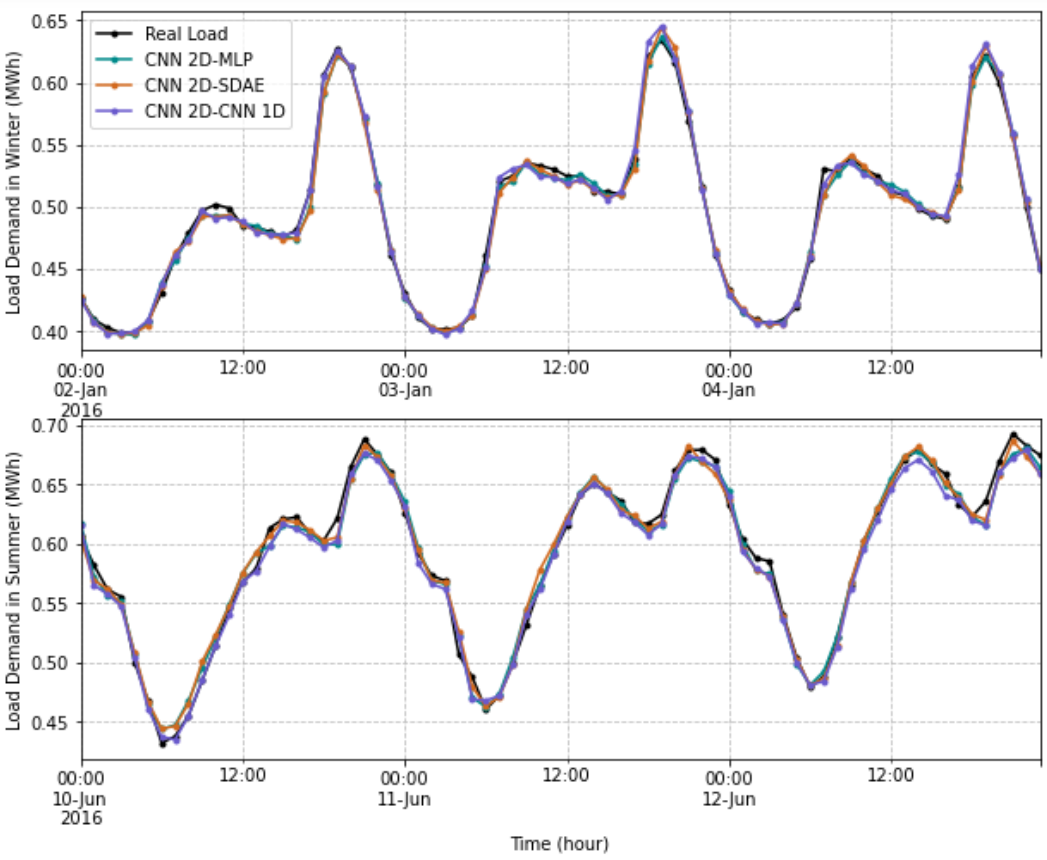


Figure 6.6: Comparison Between the Real Load Values and the Estimated Load Values of 3 Random Days in January and June Obtained Using the 3 Architectures; CNN-2D-MLP, CNN-2D-SDAE and CNN-2D-CNN-1D.

In Figure 6.6, a comparison is presented between the actual and estimated load curves across three consecutive random days during both winter and summer, employing multi-modal DL architectures such as CNN-2D-MLP, CNN-2D-SDAE, and CNN-2D-CNN-1D. The curves indicate that the overall performance is quite similar; however, when focusing on the peak loads, it becomes evident that the CNN-2D-SDAE outperformed the others. The evening peak is particularly significant for forecasting, as it signifies maximum production, and any error in this prediction could have substantial economic consequences.

6.4 Conclusion

In the current study, the STLF for Algeria was addressed utilizing multi-modal data alongside DL techniques, specifically CNN-1D, CNN-2D, SDAE and MLP.

This study was conducted using multi-modal data, requiring the development of complex and robust frameworks in the form of ANN fusion models. The multi-modal data introduced to the models is designed by the load factor, temperature factor, type of the day and the image of the past load curve to enhance the accuracy.

The obtained results emphasize both the efficiency and precision of the DANN while also illustrating the influence of multi-modal fusion within DL algorithms.

Chapter 7

Multi-level Data Fusion Based on Deep Learning

7.1 Introduction

ANN represent a powerful DL algorithms specifically designed to deliver highly accurate predictions. Significant advancements have been achieved in the development of these algorithms, enabling them to effectively process individual modalities. Their efficiency has been demonstrated across a variety of tasks. Therefore, it is essential to explore how to leverage the strengths of existing methods when they are combined, in order to achieve optimal results while handling information from multiple modalities or diverse sources of data. The premise is that if the response to each information source, whether assessed through a single DL model, is either weak or strong, then the integration of multiple models for the same task presents a promising opportunity to improve accuracy.

7.2 Data

The data is provided by the System Operator of Algerian National Electricity and Gas Company (SONELGAZ). The factors selected for the models' implementation are the ones related to the temperature and electrical load.

7.2.1 Temperature Factor

Meteorological factors and more specifically the temperature is considered to be one of the most significant external factor in STLF. The temperature related variables selected for this study encompass the maximum and minimum temperature of the day, the average temperature of the season, the hourly temperature, the temperature of the previous 24 hours, and the temperature of the 24 hours of the same day in the previous week.

7.2.2 Load Factor

In multivariate time-series models, an endogenous variable, such as the maximum national load demand, can be explained by its auto-regressive variables. The auto-regressive variables related to the load factor are: the load consumption of the previous 24 hours and the load consumption of the previous 24 hours of the same day in the previous week.

7.3 Methodologies

The fusion of information involve combining data or features from multiple modalities or sources to create a more comprehensive representation of the data. This fusion can occur at different levels of abstraction, ranging from raw data fusion to feature-level fusion, and up to decision-level fusion. In this work, the experiments were conducted to compare the different approaches based on the prediction error while executing the STLF task using SDAE and CNN-1D architectures.

7.3.1 Data-level Fusion Method

Data-level fusion, also known as early fusion or sensor-level fusion. It involves combining raw data from multiple sources or modalities into a single representation before any feature extraction or processing tasks. This approach aims to create a unified representation that captures the combined information from all sources, which can then be used for further analysis or processing using DL algorithms.

Stacked Denoising Auto-encoder Model

The proposed SDAE model consists of two DAEs. Each one is composed of an encoder and a decoder layer. The noise level was fixed to 0.5. The encoding dimensions are 60 and 40 respectively for each of the DAEs. At the fine-tuning stage, the encoding parts of the DAEs were stacked and an output layer was added at the top of the model.

One Dimensional Convolutional Neural Network Model

The CNN-1D on the other hand is composed of three convolutional layers each of which is followed by a pooling layer. The kernel size of each convolution layer is 64, 32 and 16 respectively. In the fine-tuning stage, the obtained feature map is followed by a three fully connected layers with 20 and 8 neurons in the hidden layers respectively and one neuron for the output layer.

7.3.2 Feature-level Fusion Method

The feature-level fusion based architecture processes the time-series data sources separately through the combination of the different DANNs in parallel. The meteorological factors are passed to the SDAE sub-model while the load related factors are passed to the CNN-1D sub-model for feature extraction. Then the resultant high-level features are merged into a single vector, the latter is passed into another FC meta-model to finish the prediction task and output the final decision.

The CNN-1D sub-model performs the features extraction by stacking three convolutional layers with number of kernels in each layer equals to 64, 32 and 16 respectively. The feature-map of each convolutional layer is reduced to the half by applying a max-pooling layer.

The SDAE sub-model is designed by stacking three Auto-encoders with a 0.5 added noise at each level, where the encoding dimensions are 40, 30 and 15 respectively. Each DAE is composed of an encoder and a decoder layer. After training each DAE separately, they are stacked at the fine-tuning stage and the whole model is trained as would be trained an MLP.

The high level features of the previous sub-models are merged into a single vector. A fully connected layers are added at the top of the obtained vector with 40 and 20 neurons respectively in each layer. Finally, an output

layer with 1 neuron is added at the top of the meta-model to obtain the final result.

7.3.3 Decision-level Fusion Method

Decision-level fusion frameworks are designed in a way that every sub-model has its own decision. Each sub-model deals with one data source separately without any interaction between the sub-models. The SDAE processes the meteorological data while the CNN-1D processes the load related data. Without the requirement of a meta-model, each sub-model provides a final output, then the multiple outputs are passed through a voting stage to provide the final decision.

The CNN-1D architecture consists of three convolutional layers with filter sizes 64, 32 and 16 respectively. Each layer is followed by a max-pooling layer to reduce the size of the feature-map to the half. The final feature-map is flattened to pass to the fine-tuning stage where four fully connected layers are stacked with number of neurons 20, 10, 4 and 1 respectively.

The SDAE architecture includes two DAEs with 30 and 20 encoding dimensions, respectively. Each DAE consists of an encoder and a decoder layer where the noise added to each layer is fixed to 0.5. The fine-tuning stage consists of three fully connected layers.

Figure 7.1 represents the different levels of fusion.

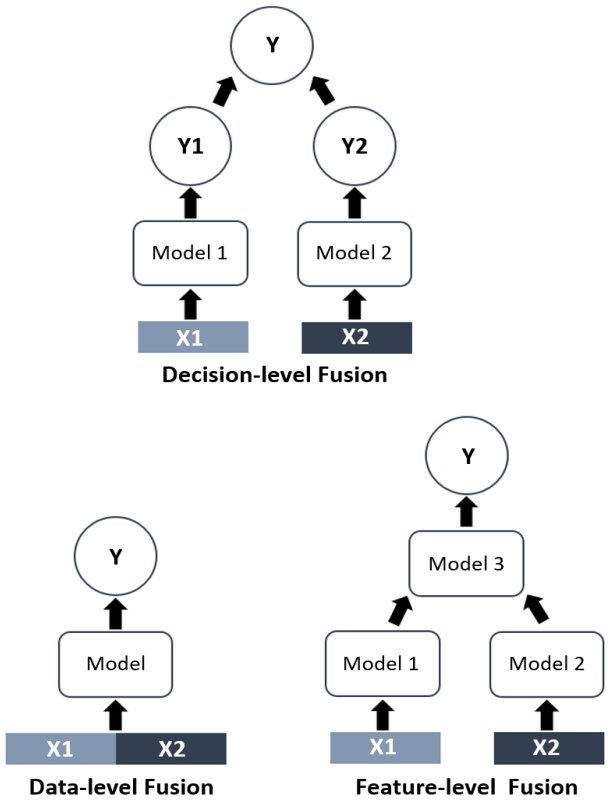


Figure 7.1: Different Levels of Fusions.

7.4 experimental results and discussion

The architectures SDAE and CNN-1D were integrated in the modeling of multi-sourced time-series data based on different levels of fusion while tackling the STLF task. The results of the different case studies displayed in Table 7.1 below are based on the MAPE performance metric Eq. (7.1) to obtain a comprehensive evaluation of the different schemes.

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{y - \hat{y}}{y} \right| \tag{7.1}$$

Where y is the real output, $\hat{y}$ is the predicted output and n is the total number of observations.

	Data-level fusion		Feature-level fusion	Decision-level fusion		
	SDAE	*CNN-1D*		*MV*	*AV*	*AVC*
MAPE (%)	1.23	0.99	0.94	1.25	4.22	2.04

Table 7.1: Summary of the results obtained with the different fusion levels based on the MAPE values.

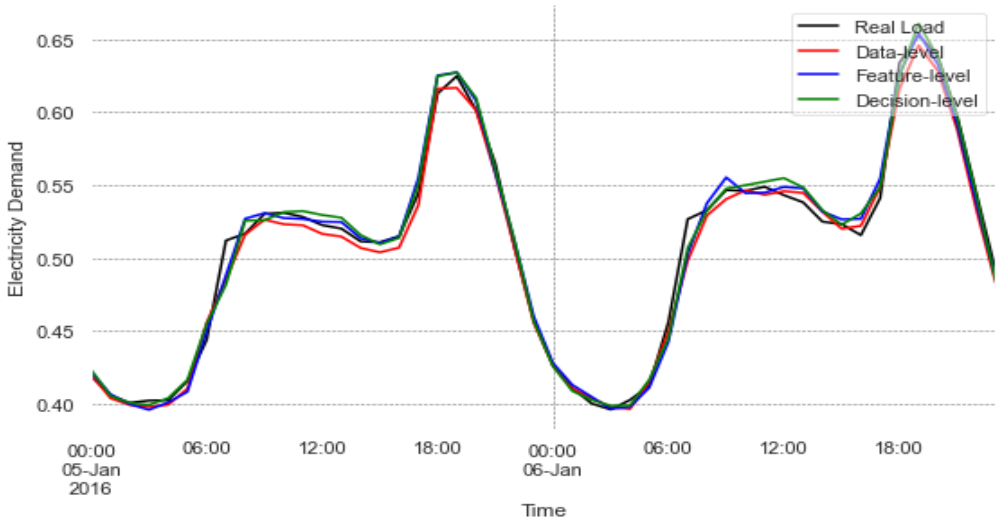


Figure 7.2: Electrical Load Demand Predicted Compared to the Original Load Profile During a Random Day in Winter.

When considering data-level fusion, the CNN-1D performed better than the SDAE model, the MAPE values are 0.99% and 1.23% respectively.

While the method based on feature-level fusion performed the best compared to the other techniques with a MAPE value of 0.94%.

On the other hand, the decision-level fusion has been categorized into three ways of voting for the final decision; Maximum voting, average voting and average voting with coefficient.

Maximum voting (MV): Simply picks the maximum output between the two final output vectors of the two models, the best MAPE value selected is 1.25%.

Average voting (AV): Is calculated by measuring the mean of the final decisions given by the two models where the MAPE value is 4.22%.

Average voting with coefficient (AVC): Gives a coefficient equals to 3 to the CNN-1D since it performed better compared to the SDAE with a MAPE value of 1.12%, while the SDAE MAPE is 8.77% as it only performs well when the input vector is large seeing the case of the data-level fusion. The MAPE obtained with the average voting with coefficient is 2.04% where the poor performance of the SDAE decreased the accuracy of the whole ensemble model.

According to the obtained results illustrated previously in Table 7.1 and Figure 7.2, we can conclude that the feature-level fusion technique outperformed the benchmark approaches. More specifically overall the CNN-1D architecture outperformed the SDAE.

7.5 Conclusion

In the present work, SDAE and CNN-1D architectures collaborated to perform the STLF of Algeria, based on data-level, feature-level and decision-level fusion of the multi-sourced information represented by the load related data and the meteorological related data provided by SONELGAZ. Furthermore, the obtained results not only underline the efficiency and precision of DANN architectures but also show the impact of the fusion level of data.

General Conclusion and Perspectives

Since AI is on a quest to imitate the human brain based on ANNs, it's imperative that it processes multi-modal information simultaneously. Overall, the integration of multi-modal AI holds great promise for advancing the capabilities of ANNs in processing multi-modal information and achieving SOTA performance in various real-world applications In recent years, we witnessed the emergence of AI adoption by the companies such as electrical companies to perform several tasks, since it is the only successful way to alleviate risks. In particular; reducing costs, boosting revenues and saving time. One of the most common tasks is the prediction of electricity demand, this topic has been an important issue for several decades. Yet, there is still much progress to be made in this field. This thesis is concerned with multi-modal AI methodologies and its application to STLF. In particular, we shed the light on different tasks related to electricity demand forecasting. The first contribution is the improvement of the STLF models based on auto-regressive variables using one modality of data. The second contribution is the introduction multi-modal information in the process of STLF. The last contribution was to perform data fusion based on DL techniques using the available data provided by SONELGAZ. The obtained results not only underline the efficiency and precision of the DANN models, but also show the impact of multi-modality in DL. The clear performance gain over the uni-modal AI underlines the effectiveness and precision of multi-modal fusion. Multi-modal AI represents a forward step in how developers apply the functionality of AI to the next generation of applications. In future research, we look forward to explore other aspects of multi-modal AI and expand the multi-modal applications to generative AI.

Scientific Productions

Chelabi, H., Khadir, M. T., Chikhaoui, B. (2022). Deep neural network architectures for electrical load forecasting: A review. Int. J. of Artificial Intelligence Tools and Application (AITA), 1(1), 1-22.

Hiba, C., Tarek, K. M., Belkacem, C. (2020, June). Stacked Denoising Autoencoder network for short-term prediction of electrical Algerian load. In 2020 7th International Conference on Control, Decision and Information Technologies (CoDIT) (Vol. 1, pp. 189-194). IEEE.
DOI: 10.1109/CoDIT49905.2020.9263850

H. Chelabi, M. T. Khadir, B. Chikhaoui A. J. Telmoudi (2022) Comparison of Deep Learning Architectures for Short-Term Electrical Load Forecasting Based on Multi-Modal Data, Cybernetics and Systems, 53:1, 186-207, DOI: 10.1080/01969722.2021.2008679

C. Hiba, K. M. Tarek and C. Belkacem, "Multi-Level Fusion of Multi-Source Information Based Deep Learning and Ensemble Deep Learning Models," 2023 9th International Conference on Control, Decision and Information Technologies (CoDIT), Rome, Italy, 2023, pp. 315-320,
doi: 10.1109/CoDIT58514.2023.10284131.

Bibliography

[1] Russell, T. (2014). Multimodal representations and science learning. In Encyclopedia of Science Education (pp. 1-8). Springer, Dordrecht.

[2] Kong, Z., Zhang, C., Lv, H., Xiong, F., Fu, Z. (2020). Multimodal feature extraction and fusion deep neural networks for short-term load forecasting. IEEE access, 8, 185373-185383.

[3] Xian, Q., Liang, W. (2021, December). A multi-modal time series intelligent prediction model. In INTERNATIONAL CONFERENCE ON WIRELESS COMMUNICATIONS, NETWORKING AND APPLICATIONS (pp. 1150-1157). Singapore: Springer Nature Singapore.

[4] Trinugroho, I., Susilowati, F. (2022). Simultaneous analysis: The effect of electricity consumption on human development index in asean 5. JEJAK: Jurnal Ekonomi dan Kebijakan, 15(2), 234-243.

[5] Sarkodie, S. A., & Adams, S. (2020). Electricity access, human development index, governance and income inequality in Sub-Saharan Africa. Energy Reports, 6, 455-466.

[6] McCarthy, J., Minsky, M. L., Rochester, N., Shannon, C. E. (2006). A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. AI magazine, 27(4), 12-12.

[7] Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y. (2016). Deep Learning Cambridge. MA: MIT Press http://www. deeplearningbook. org.

[8] Rosenblatt, F. (2007). Perceptron simulation experiments. Proceedings of the IRE, 48(3), 301-309.

[9] Rumelhart, D. E., Hinton, G. E., Williams, R. J. (1986). Learning representations by back-propagating errors. nature, 323(6088), 533-536.

[10] Haykin, S. (2009). Neural networks and learning machines, 3/E. Pearson Education India.

[11] Baldi, P. (2012, June). Autoencoders, unsupervised learning, and deep architectures. In Proceedings of ICML workshop on unsupervised and transfer learning (pp. 37-49). JMLR Workshop and Conference Proceedings.

[12] Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. science, 313(5786), 504-507.

[13] Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P. A., Bottou, L. (2010). Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. Journal of machine learning research, 11(12).

[14] Bishop, C. M. (1995). Training with noise is equivalent to Tikhonov regularization. Neural computation, 7(1), 108-116.

[15] Larochelle, H., Erhan, D., Courville, A., Bergstra, J., & Bengio, Y. (2007, June). An empirical evaluation of deep architectures on problems with many factors of variation. In Proceedings of the 24th international conference on Machine learning (pp. 473-480).

[16] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. Neural computation, 1(4), 541-551.

[17] Srinivasamurthy, R. S. (2018). Understanding 1D convolutional neural networks using multiclass time-varying signalss (Doctoral dissertation, Clemson University).

[18] Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., & Inman, D. J. (2021). 1D convolutional neural networks and applications: A survey. Mechanical Systems and Signal Processing, 151, 107398.

[19] Bernsen, N. O., Dybkjær, L. (2009). Multimodal usability. Springer Science & Business Media.

[20] Rahate, A., Walambe, R., Ramanna, S., Kotecha, K. (2022). Multimodal co-learning: Challenges, applications with data-sets, recent advances and future directions. Information Fusion, 81, 203-239.

[21] Kress, G. (2009). Multimodality: A social semiotic approach to contemporary communication. routledge.

[22] Baltrušaitis, T., Ahuja, C., & Morency, L. P. (2018). Multimodal machine learning: A survey and taxonomy. IEEE transactions on pattern analysis and machine intelligence, 41(2), 423-443.

[23] Ito, F. T., de Medeiros Caseli, H., Moreira, J. (2018, May). The effects of unimodal representation choices on multimodal learning. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018).

[24] Zhang, S. F., Zhai, J. H., Xie, B. J., Zhan, Y., & Wang, X. (2019, July). Multimodal representation learning: Advances, trends and challenges. In 2019 International Conference on Machine Learning and Cybernetics (ICMLC) (pp. 1-6). IEEE.

[25] Onyshchak, O. (2020). Image Recommendation for Wikipedia Articles.

[26] Liang, P. P., Zadeh, A., Morency, L. P. (2022). Foundations and Trends in Multimodal Machine Learning: Principles, Challenges, and Open Questions. arXiv preprint arXiv:2209.03430.

[27] S. -F. Zhang, J. -H. Zhai, B. -J. Xie, Y. Zhan and X. Wang, "Multimodal Representation Learning: Advances, Trends and Challenges," 2019 International Conference on Machine Learning and Cybernetics (ICMLC), Kobe, Japan, 2019, pp. 1-6, doi: 10.1109/ICMLC48188.2019.8949228. keywords: Multimodal;Representation learning;Machine learning;deep learning;Multimodal deep learning,

[28] Aeberhard, M., Kaempchen, N. (2011, March). High-level sensor data fusion architecture for vehicle surround environment perception. In Proc. 8th Int. Workshop Intell. Transp (Vol. 665, pp. 1-7).

[29] Al-bayati, J. S. H., Üstündağ, B. B. (2020). Early and Late Fusion of Deep Convolutional Neural Networks and Evolutionary feature optimization for Plant leaf illness recognition. JX Univ. of Archit. Technol., 12(2), 1591-1610.

[30] Shi, T., Xu, Q., Zou, Z., Shi, Z. (2018). Automatic raft labeling for remote sensing images via dual-scale homogeneous convolutional neural network. Remote Sensing, 10(7), 1130.

[31] Pandeya, Y. R., Lee, J. (2021). Deep learning-based late fusion of multimodal information for emotion classification of music video. Multimedia Tools and Applications, 80(2), 2887-2905.

[32] Ahsan, H., Bhalla, N., Bhatt, D., Shah, K. (2021). Multi-modal image captioning for the visually impaired. arXiv preprint arXiv:2105.08106.

[33] Nguyen, T., Gadre, S. Y., Ilharco, G., Oh, S., Schmidt, L. (2024). Improving multimodal data-sets with image captioning. Advances in Neural Information Processing Systems, 36.

[34] Yu, J., Li, J., Yu, Z., Huang, Q. (2019). Multimodal transformer with multi-view visual representation for image captioning. IEEE transactions on circuits and systems for video technology, 30(12), 4467-4480.

[35] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

[36] Zhao, D., Chang, Z., Guo, S. (2019). A multimodal fusion approach for image captioning. Neurocomputing, 329, 476-485.

[37] Kuo, C. W., Kira, Z. (2022). Beyond a pre-trained object detector: Cross-modal textual and visual context for image captioning. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 17969-17979).

[38] Tian, Y., Newsam, S., Boakye, K. (2023). Fashion image retrieval with text feedback by additive attention compositional learning. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 1011-1021).

[39] He, L., Liu, S., An, R., Zhuo, Y., Tao, J. (2023). An end-to-end framework based on vision-language fusion for remote sensing cross-modal text-image retrieval. Mathematics, 11(10), 2279.

[40] Nguyen, T., Hendriksen, M., Yates, A. (2024). Multimodal Learned Sparse Retrieval for Image Suggestion. arXiv preprint arXiv:2402.07736.

[41] Dhiman, G., Kumar, A. V., Nirmalan, R., Sujitha, S., Srihari, K., Yuvaraj, N., ... Raja, R. A. (2023). Multi-modal active learning with deep reinforcement learning for target feature extraction in multi-media image processing applications. Multimedia Tools and Applications, 82(4), 5343-5367.

[42] Liu, Z., Sun, W., Hong, Y., Teney, D., Gould, S. (2024). Bi-directional training for composed image retrieval via text prompt learning. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 5753-5762).

[43] Xu, Y., Zhang, L., Shen, X. (2023). Multi-modal adaptive gated mechanism for visual question answering. Plos one, 18(6), e0287557.

[44] Guo, Z., Han, D. (2023). Multi-modal co-attention relation networks for visual question answering. The Visual Computer, 39(11), 5783-5795.

[45] Xu, Y., Zhang, L., Shen, X. (2023). Multi-modal adaptive gated mechanism for visual question answering. Plos one, 18(6), e0287557.

[46] Wang, Y., Yasunaga, M., Ren, H., Wada, S., Leskovec, J. (2023). Vqa-gnn: Reasoning with multimodal knowledge via graph neural networks for visual question answering. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 21582-21592).

[47] Lyu, C., Li, W., Ji, T., Zhou, L., Gurrin, C. (2023). Gated Multimodal Fusion with Cross-modal Contrastive Learning for Video Question Answering. In: Iliadis, L., Papaleonidas, A., Angelov, P., Jayne, C. (eds) Artificial Neural Networks and Machine Learning – ICANN 2023. ICANN 2023. Lecture Notes in Computer Science, vol 14260. Springer, Cham.

[48] Qian, T., Chen, J., Zhuo, L., Jiao, Y., Jiang, Y. G. (2023). Nuscenes-qa: A multi-modal visual question answering benchmark for autonomous driving scenario. arXiv preprint arXiv:2305.14836.

[49] Lian H, Lu C, Li S, Zhao Y, Tang C, Zong Y. A Survey of Deep Learning-Based Multimodal Emotion Recognition: Speech, Text, and Face. Entropy. 2023; 25(10):1440. https://doi.org/10.3390/e25101440

[50] Jun Sun, Shoukang Han, Yu-Ping Ruan, Xiaoning Zhang, Shu-Kai Zheng, Yulong Liu, Yuxin Huang, and Taihao Li. 2023. Layer-wise Fusion with Modality Independence Modeling for Multi-modal Emotion Recognition. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 658–670, Toronto, Canada. Association for Computational Linguistics.

[51] Wu, Y., Li, J. Multi-modal emotion identification fusing facial expression and EEG. Multimed Tools Appl 82, 10901–10919 (2023). https://doi.org/10.1007/s11042-022-13711-4

[52] Dempster, A. (1967). Upper and lower probabilities induced by multivalued mapping, A. of Mathematical Statistics, Ed. AMS-38, 10.

[53] Shafer, G. (1976). A mathematical theory of evidence (Vol. 42). Princeton university press.

[54] Dempster, A.P. (2008). Upper and Lower Probabilities Induced by a Multivalued Mapping. In: Yager, R.R., Liu, L. (eds) Classic Works of the Dempster-Shafer Theory of Belief Functions. Studies in Fuzziness and Soft Computing, vol 219. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-44792-4$_3$

[55] Ge S, Jiang Z, Cheng Z, Wang C, Yin Y, Gu Q (2023) Learning robust multi-modal representation for multi-label emotion recognition via adversarial masking and perturbation. In: Proceedings of the ACM Web Conference 2023. WWW '23, Association for Computing Machinery, New York, USA, pp 1510–1518. https://doi.org/10.1145/3543507.3583258

[56] Dutta, S., Ganapathy, S. (2023). HCAM–Hierarchical Cross Attention Model for Multi-modal Emotion Recognition. arXiv preprint arXiv:2304.06910.

[57] Adel, O., Fathalla, K. M., Abo ElFarag, A. (2023). MM-EMOR: Multi-Modal Emotion Recognition of Social Media Using Concatenated Deep Learning Networks. Big Data and Cognitive Computing, 7(4), 164.

[58] Xue, J., Deng, Y., Wang, F., Li, Y., Gao, Y., Tao, J., ... Liang, J. (2023, June). M 2-ctts: End-to-end multi-scale multi-modal conversational text-to-speech synthesis. In ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1-5). IEEE.

[59] Kim, M., Choi, J., Maiti, S., Yeo, J. H., Watanabe, S., Ro, Y. M. (2023). Towards practical and efficient image-to-speech captioning with vision-language pre-training and multi-modal tokens. arXiv preprint arXiv:2309.08531.

[60] Kim, M., Jung, J. W., Rha, H., Maiti, S., Arora, S., Chang, X., ... Ro, Y. M. (2024). TMT: Tri-Modal Translation between Speech, Image, and Text by Processing Different Modalities as Different Languages. arXiv preprint arXiv:2402.16021.

[61] Huang, R., Li, M., Yang, D., Shi, J., Chang, X., Ye, Z., ... Watanabe, S. (2023). Audiogpt: Understanding and generating speech, music, sound, and talking head. arXiv preprint arXiv:2304.12995.

[62] Guan, W., Li, Y., Li, T., Huang, H., Wang, F., Lin, J., ... Hong, Q. (2023). MM-TTS: Multi-modal Prompt based Style Transfer for Expressive Text-to-Speech Synthesis. arXiv preprint arXiv:2312.10687.

[63] Liu, M., Liang, K., Hu, D., Yu, H., Liu, Y., Meng, L., ... Liu, X. (2023, October). Tmac: Temporal multi-modal graph learning for acoustic event classification. In Proceedings of the 31st ACM International Conference on Multimedia (pp. 3365-3374).

[64] Wang, S., Ju, M., Zhang, Y., Zheng, Y., Wang, M., Qi, G. (2023, April). Cross-modal contrastive learning for event extraction. In International Conference on Database Systems for Advanced Applications (pp. 699-715). Cham: Springer Nature Switzerland.

[65] Fayyaz, H., Strang, A., Beheshti, R. (2023, December). Bringing at-home pediatric sleep apnea testing closer to reality: A multi-modal transformer approach. In Machine Learning for Healthcare Conference (pp. 167-185). PMLR.

[66] Wu, L., Liu, P., Zhao, Y., Wang, P., Zhang, Y. (2023). Human cognition-based consistency inference networks for multi-modal fake news detection. IEEE Transactions on Knowledge and Data Engineering.

[67] Li, W., Liu, X., Wang, D., Lu, W., Yuan, B., Qin, C., ... Căleanu, C. (2024). MITDCNN: A multi-modal input Transformer-based deep convolutional neural network for misfire signal detection in high-noise diesel engines. Expert Systems with Applications, 238, 121797.

[68] Korban, M., Youngs, P., Acton, S. T. (2023). A Multi-Modal Transformer network for action detection. Pattern Recognition, 142, 109713.

[69] Lee, S., Woo, S., Park, Y., Nugroho, M. A., Kim, C. (2023). Modality mixer for multi-modal action recognition. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 3298-3307).

[70] Chakraborty, A., Mukherjee, N. (2023). A deep-CNN based low-cost, multi-modal sensing system for efficient walking activity identification. Multimedia Tools and Applications, 82(11), 16741-16766.

[71] Yani, M., Yamada, N., Siow, C. Z., Kubota, N. (2023). An efficient activity recognition for homecare robots from multi-modal communication data-set. International Journal of Advances in Intelligent Informatics, 9(1).

[72] Woo, S., Lee, S., Park, Y., Nugroho, M. A., Kim, C. (2023, June). Towards good practices for missing modality robust action recognition. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 37, No. 3, pp. 2776-2784).

[73] Li, Y., Yu, Z., Xiang, S., Liu, T., Fu, Y. (2023, June). Av-tad: Audio-visual temporal action detection with transformer. In ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1-5). IEEE.

[74] Georgilakis, P. S. (2009). Spotlight on modern transformer design. Springer Science Business Media.

[75] Nasir, N., Kansal, A., Barneih, F., Al-Shaltone, O., Bonny, T., Al-Shabi, M., Al Shammaa, A. (2023). Multi-modal image classification of COVID-19 cases using computed tomography and X-rays scans. Intelligent Systems with Applications, 17, 200160.

[76] Xing, Z., He, Y. (2023). Multi-modal multi-step wind power forecasting based on stacking deep learning model. Renewable Energy, 215, 118991.

[77] Lv, T., Huang, Y., Chen, J., Cui, L., Ma, S., Chang, Y., ... Wei, F. (2023). Kosmos-2.5: A multimodal literate model. arXiv preprint arXiv:2309.11419.

[78] Mia, M. S., Tanabe, R., Habibi, L. N., Hashimoto, N., Homma, K., Maki, M., ... Tanaka, T. S. (2023). Multimodal deep learning for rice yield prediction using UAV-based multispectral imagery and weather data. Remote Sensing, 15(10), 2511.

[79] Kolluri, J., Das, R. (2023). Intelligent multimodal pedestrian detection using hybrid metaheuristic optimization with deep learning model. Image and Vision Computing, 131, 104628.

[80] Behm, C., Nolting, L., & Praktiknjo, A. (2020). How to model European electricity load profiles using artificial neural networks. Applied Energy, 277, 115564.

[81] Wang, M., Zheng, Y., Wang, B., & Deng, Z. (2021). Household Electricity Load Forecasting Based on Multitask Convolutional Neural Network with Profile Encoding. Mathematical Problems in Engineering, 2021.

[82] Rafi, S. H., Deeba, S. R., & Hossain, E. (2021). A Short-Term Load Forecasting Method Using Integrated CNN and LSTM Network. IEEE Access, 9, 32436-32448.

[83] Mishra, K., Basu, S., & Maulik, U. (2021). A Dilated Convolutional Based Model for Time Series Forecasting. SN Computer Science, 2(2), 1-11.

[84] Aurangzeb, K., Alhussein, M., Javaid, K., & Haider, S. I. (2021). A Pyramid-CNN Based Deep Learning Model for Power Load Forecasting of Similar-Profile Energy Customers Based on Clustering. IEEE Access, 9, 14992-15003.

[85] Xuan, Y., Si, W., Zhu, J., Sun, Z., Zhao, J., Xu, M., & Xu, S. (2021). Multi-model fusion short-term load forecasting based on random forest feature selection and hybrid neural network. IEEE Access.

[86] Wang, J., Chen, X., Zhang, F., Chen, F., & Xin, Y. (2021). Building load forecasting using deep neural network with efficient feature fusion. Journal of Modern Power Systems and Clean Energy, 9(1), 160-169.

[87] Shen, Y., Ma, Y., Deng, S., Huang, C. J., & Kuo, P. H. (2021). An ensemble model based on deep learning and data preprocessing for short-term electrical load forecasting. Sustainability, 13(4), 1694.

[88] Andriopoulos, N., Magklaras, A., Birbas, A., Papalexopoulos, A., Valouxis, C., Daskalaki, S., ... & Papaioannou, G. P. (2021). short-term Electric Load Forecasting Based on Data Transformation and Statistical Machine Learning. Applied Sciences, 11(1), 158.

[89] Wang, R., & Zhao, J. (2020). Deep Learning-Based Short-Term Load Forecasting for Transformers in Distribution Grid. International Journal of Computational Intelligence Systems.

[90] Tudose, A. M., Sidea, D. O., Picioroaga, I. I., Boicea, V. A., & Bulac, C. (2020, September). A CNN Based Model for Short-Term Load Forecasting: A Real Case Study on the Romanian Power System. In 2020 55th International Universities Power Engineering Conference (UPEC) (pp. 1-6). IEEE.

[91] Peng, Q., & Liu, Z. W. (2020, July). Short-Term Residential Load Forecasting Based on Smart Meter Data Using Temporal Convolutional Networks. In 2020 39th Chinese Control Conference (CCC) (pp. 5423-5428). IEEE.

[92] Xie, M., Chai, C., Guo, H., & Wang, M. (2020, July). Household Electricity Load Forecasting Based on Pearson Correlation Coefficient Clustering and Convolutional Neural Network. In Journal of Physics: Conference Series (Vol. 1601, No. 2, p. 022012). IOP Publishing.

[93] Huang, Q., Li, J., & Zhu, M. (2020). An improved convolutional neural network with load range discretization for probabilistic load forecasting. Energy, 203, 117902.

[94] Khan, I. U., Javaid, N., Taylor, C. J., Gamage, K. A., & Ma, X. (2020, July). Big Data Analytics Based Short Term Load Forecasting Model for Residential Buildings in Smart Grids. In IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS) (pp. 544-549). IEEE.

[95] Moradzadeh, A., Moayyed, H., Zakeri, S., Mohammadi-Ivatloo, B., & Aguiar, A. P. (2021). Deep Learning-Assisted Short-Term Load Forecasting for Sustainable Management of Energy in Microgrid. Inventions, 6(1), 15.

[96] Farsi, B., Amayri, M., Bouguila, N., & Eicker, U. (2021). On short-term load forecasting using machine learning techniques and a novel parallel deep LSTM-CNN approach. IEEE Access, 9, 31191-31212.

[97] Mathew, J., & Behera, R. K. (2021). Power load forecasting based on long short-term memory-singular spectrum analysis. Energy Systems, 1-23.

[98] Ayub, N., Irfan, M., Awais, M., Ali, U., Ali, T., Hamdi, M., ... & Muhammad, F. (2020). Big Data Analytics for Short and Medium-Term Electricity Load Forecasting Using an AI Techniques Ensembler. Energies, 13(19), 5193.

[99] Jin, X. B., Zheng, W. Z., Kong, J. L., Wang, X. Y., Bai, Y. T., Su, T. L., & Lin, S. (2021). Deep-Learning Forecasting Method for Electric Power Load via Attention-Based Encoder-Decoder with Bayesian Optimization. Energies, 14(6), 1596.

[100] Ullah, F. U. M., Khan, N., Hussain, T., Lee, M. Y., & Baik, S. W. (2021). Diving Deep into Short-Term Electricity Load Forecasting: Comparative Analysis and a Novel Framework. Mathematics, 9(6), 611.

[101] Adewuyi, S. A., Aina, S., & Oluwaranti, A. I. (2020). A deep learning model for electricity demand forecasting based on a tropical data. Applied Computer Science, 16(1).

[102] Memarzadeh, G., & Keynia, F. (2021). Short-term electricity load and price forecasting by a new optimal LSTM-NN based prediction algorithm. Electric Power Systems Research, 192, 106995.

[103] ALTUNKAYA, D., & YILMAZ, B. Multivariate Short-term Load Forecasting Using Deep Learning Algorithms. The Eurasia Proceedings of Science Technology Engineering and Mathematics, 11, 14-19.

[104] Mahdy, B., Abbas, H., Hassanein, H. S., Noureldin, A., & Abou-zeid, H. (2020). A Clustering-Driven Approach to Predict the Traffic Load of Mobile Networks for the Analysis of Base Stations Deployment. Journal of Sensor and Actuator Networks, 9(4), 53.

[105] Fekri, M. N., Patel, H., Grolinger, K., & Sharma, V. (2021). Deep learning for load forecasting with smart meter data: Online Adaptive Recurrent Neural Network. Applied Energy, 282, 116177.

[106] Suhagiya, J., Raval, D., Pandey, S. V., Patel, J., Gupta, A., & Srivastava, A. (2020). RECURRENT NEURAL NETWORK BASED ELECTRICITY LOAD FORECASTING OF G-20 MEMBERS.

[107] Xuan, W., Shouxiang, W., Qianyu, Z., Shaomin, W., & Liwei, F. (2021). A multi-energy load prediction model based on deep multi-task learning and ensemble approach for regional integrated energy systems. International Journal of Electrical Power & Energy Systems, 126, 106583.

[108] Massaoudi, M., Refaat, S. S., Chihi, I., Trabelsi, M., Abu-Rub, H., & Oueslati, F. S. (2020, October). Short-Term Electric Load Forecasting Based on Data-Driven Deep Learning Techniques. In IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society (pp. 2565-2570). IEEE.

[109] He, H., Wang, H., Ma, H., Liu, X., Jia, Y., & Gong, G. (2020, October). Research on Short-term Power Load Forecasting Based on Bi-GRU. In Journal of Physics: Conference Series (Vol. 1639, No. 1, p. 012017). IOP Publishing.

[110] Hsu, C. C., Chen, X. T., Chen, Y. S., & Chang, A. (2020, September). Short-Term Load Forecasting by Machine Learning. In 2020 International Symposium on Community-centric Systems (CcS) (pp. 1-6). IEEE.

[111] Liu, T., Zhang, Y., Zhao, H., Liu, X., Gao, T., Yuan, H., & Zhang, J. (2020, August). Social Implications of Cyber-Physical Systems in Electrical Load Forecasting. In 2020 IEEE 16th International Conference on Automation Science and Engineering (CASE) (pp. 582-587). IEEE.

[112] Mishra, S., Petersen, A., Frank, S. M., & Slovensky, M. (2020). Deep Learning Based Load Forecasting: from Research to Deployment–Opportunities and Challenges. arXiv preprint arXiv:2008.05458.

[113] Zhang, L., Yang, L., Gu, C., & Li, D. (2020). LSTM-based Short-term Electrical Load Forecasting and Anomaly Correction. In E3S Web of Conferences (Vol. 182, p. 01004). EDP Sciences.

[114] Pang, C., Bao, T., & He, L. (2020). Power System Load Forecasting Method Based on Recurrent Neural Network. In E3S Web of Conferences (Vol. 182, p. 02007). EDP Sciences.

[115] Chen, L., Yu, H., Tong, L., Huai, X., Jin, P., Huang, Y., & Dou, C. (2020, August). Research on Load Forecasting Method of Distribution Transformer based on Deep Learning. In 2020 7th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2020 6th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom) (pp. 228-233). IEEE.

[116] Yang, Y., Haq, E. U., & Jia, Y. (2020, July). A Novel Deep Learning Approach for Short and Medium-Term Electrical Load Forecasting Based on Pooling LSTM-CNN Model. In 2020 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia) (pp. 26-34). IEEE.

[117] Cini, A., Lukovic, S., & Alippi, C. (2020, July). Cluster-based Aggregate Load Forecasting with Deep Neural Networks. In 2020 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.

[118] He, Y., Henze, J., & Sick, B. (2020, July). Forecasting power grid states for regional energy markets with deep neural networks. In 2020 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.

[119] Jahangir, H., Tayarani, H., Gougheri, S. S., Golkar, M. A., Ahmadian, A., & Elkamel, A. (2020). Deep learning-based forecasting approach in smart grids with micro-clustering and bi-directional LSTM network. IEEE Transactions on Industrial Electronics.

[120] Atef, S., & Eltawil, A. B. (2020). Assessment of stacked unidirectional and bidirectional long short-term memory networks for electricity load forecasting. Electric Power Systems Research, 187, 106489.

[121] Zhang, L., Shi, J., Wang, L., & Xu, C. (2020). Electricity, Heat, and Gas Load Forecasting Based on Deep Multitask Learning in Industrial-Park Integrated Energy System. Entropy, 22(12), 1355.

[122] Zhou, B., Meng, Y., Huang, W., Wang, H., Deng, L., Huang, S., & Wei, J. (2021). Multi-energy net load forecasting for integrated local energy systems with heterogeneous prosumers. International Journal of Electrical Power & Energy Systems, 126, 106542.

[123] Huan, J., Hong, H., Pan, X., Sui, Y., Zhang, X., Jiang, X., & Wang, C. (2020, June). Short-Term Load Forecasting of Integrated Energy Systems Based on Deep Learning. In 2020 5th Asia Conference on Power and Electrical Engineering (ACPEE) (pp. 16-20). IEEE.

[124] Tang, X., Dai, Y., Liu, Q., Dang, X., & Xu, J. (2019). Application of bidirectional recurrent neural network combined with deep belief network in short-term load forecasting. IEEE Access, 7, 160660-160670.

[125] Xu, B., Wu, T., Wang, X., Zhu, Y., Guo, N., & Cai, Z. (2021, March). Research on load forecasting method of large Power Grid based on Deep confidence Network. In IOP Conference Series: Earth and Environmental Science (Vol. 692, No. 2, p. 022117). IOP Publishing.

[126] Dong, Y., Ma, X., & Fu, T. (2021). Electrical load forecasting: A deep learning approach based on K-nearest neighbors. Applied Soft Computing, 99, 106900.

[127] Na, Z., HanZhen, T., YuTong, L., Jia, C., JunYou, Y., & Wang, G. (2020, July). Short-term load forecasting algorithm based on LSTM-DBN considering the flexibility of electric vehicle. In IOP Conference Series: Earth and Environmental Science (Vol. 546, No. 4, p. 042001). IOP Publishing.

[128] Dong, Y., Dong, Z., Zhao, T., Li, Z., Ding, Z. (2021). Short term load forecasting with markovian switching distributed deep belief networks. International Journal of Electrical Power Energy Systems, 130, 106942.

[129] Rong, Y., Li, J., Ju, W., Tang, X., Liu, Y., Xu, X., ... Shen, H. (2021, September). An integrated energy system load prediction study based on deep belief networks and multitasking learning. In Journal of Physics: Conference Series (Vol. 2035, No. 1, p. 012002). IOP Publishing.

[130] Wang, T., Lai, C. S., Ng, W. W., Pan, K., Zhang, M., Vaccaro, A., Lai, L. L. (2021). Deep autoencoder with localized stochastic sensitivity for short-term load forecasting. International Journal of Electrical Power Energy Systems, 130, 106954.

[131] Son, M., Moon, J., Jung, S., Hwang, E. (2018, September). A short-term load forecasting scheme based on auto-encoder and random forest. In International Conference on Applied Physics, System Science and Computers (pp. 138-144). Springer, Cham.

[132] Hamedmoghadam, H., Joorabloo, N., & Jalili, M. (2018). Australia's long-term electricity demand forecasting using deep neural networks. arXiv preprint arXiv:1801.02148.

[133] Ke, K., Hongbin, S., Chengkang, Z., & Brown, C. (2019). Short-term electrical load forecasting method based on stacked auto-encoding and GRU neural network. Evolutionary Intelligence, 12(3), 385-394.

[134] Peng, W., Xu, L., Li, C., Xie, X., & Zhang, G. (2019). Stacked autoencoders and extreme learning machine based hybrid model for electrical load prediction. Journal of Intelligent & Fuzzy Systems, 37(4), 5403-5416.

[135] Huang, X., Hu, T., Ye, C., Xu, G., Wang, X., & Chen, L. (2019). Electric load data compression and classification based on deep stacked AEs. Energies, 12(4), 653.

[136] Tong, C., Li, J., Lang, C., Kong, F., Niu, J., & Rodrigues, J. J. (2018). An efficient deep model for day-ahead electricity load forecasting with stacked denoising AEs. Journal of Parallel and Distributed Computing, 117, 267-273.

[137] Liu, P., Zheng, P., & Chen, Z. (2019). Deep learning with stacked denoising auto-encoder for short-term electric load forecasting. Energies, 12(12), 2445.

[138] Su, T., Liu, Y., Zhao, J., Liu, J. (2020). Probabilistic Stacked Denoising Autoencoder for Power System Transient Stability Prediction with Wind Farms. IEEE Transactions on Power Systems.

[139] Khadir, M. T., Farah, N., Farfar, K., Bendaoud, N., Ameyoud, A., & Tenzer, N. (2019, February). PREVOS-DZ: A Short-Mid Term Algerian Electric Load Forecasting Software. In 2019 Algerian Large Electrical Network Conference (CAGRE) (pp. 1-6). IEEE.

[140] Bendaoud, N. M. M., & Farah, N. (2020). Using deep learning for short-term load forecasting. Neural Computing and Applications, 32(18), 15029-15041.

[141] Farfar, K. E., & Khadir, M. T. (2019). A two-stage short-term load forecasting approach using temperature daily profiles estimation. Neural Computing and Applications, 31(8), 3909-3919.

[142] F. Bélaid, F. Abderrahmani, Electricity consumption and economic growth in Algeria: A multivariate causality analysis in the presence of structural change, Energy Policy, Volume 55, pp. 286-295, 2013.

[143] Moral-Carcedo, J., Pérez-García, J. (2019). Time of day effects of temperature and daylight on short-term electricity load. Energy, 174, 169-183.

[144] R. M. Nezzar, N. Farah, M. T. Khadir, L. Chouireb, Mid-Long Term Load Forecasting using Multi-Model Artificial Neural Networks, International Journal on Electrical Engineering and Informatics, Volume 8, no 2, 2016.

[145] K. E. Farfar, M. T. Khadir, O. Laid, Comparison of serial and parallel approaches using artificial neural networks for Algerian short-term load forecasting, Conf. on Advances in Computing, Electronics and Electrical Technology, 2015.

[146] E. Dale, E. Linvill, Calculating Chilling Hours and Chill Units from Daily Maximum and Minimum Temperature Observations, HORTSCIENCE, Volume 25, 1990.

[147] J. Spencer, Fourier series representation of the position of the sun, Search, Volume 2, PP. 172, 1971.

[148] J. Almorox, C. Hontoria, M. Benito, Statistical validation of daylength definitions for estimation of global solar radiation in Toledo, Spain, Energy Conversion and Management, Volume 46, pp. 1465–1471, 2005.

[149] S. C. Nayak, B. B. Misra, H. S. Behera, Impact of Data Normalization on Stock Index Forecasting, International Journal of Computer Information Systems and Industrial Management Applications, Volume 6, pp. 257-269, 2014.