



T.C.

ALTINBAŞUNIVERSITY
Institute of Graduate Studies
Information Technologies

**EFFICIENT CONSTRUCTION COST ESTIMATION
SOFTWARE USING CONVOLUTIONAL NEURAL
NETWORK AND PARTICLE SWARM
OPTIMIZATION ALGORITHM (PSO)**

Mohanad Mohammed Hamad ALSABAAWI

Master's Thesis

Supervisor

Prof. Dr. Osman Nuri UCAN

Istanbul, 2020

**EFFICIENT CONSTRUCTION COST ESTIMATION SOFTWARE USING
CONVOLUTIONAL NEURAL NETWORK AND PARTICLE SWARM
OPTIMIZATION ALGORITHM (PSO)**

by

Mohanad Mohammed Hamad ALSABAAWI

Information Technologies

Submitted to the Institute of Graduate Studies

in partial fulfillment of the requirements for the degree of

Master of Science

ALTINBAŞ UNIVERSITY

2020

The thesis titled “EFFICIENT CONSTRUCTION COST ESTIMATION SOFTWARE USING CONVOLUTIONAL NEURAL NETWORK AND PARTICLE SWARM OPTIMIZATION ALGORITHM (PSO)” prepared and presented by “MOHANAD MOHAMMED HAMAD ALSABAAWI” was accepted as a Master of Science Thesis in Information Technologies.

Prof. Dr. Osman Nuri Ucan
Supervisor

Thesis Defense Jury Members:

Prof. Dr. Osman Nuri Ucan

School of Engineering and
Natural Sciences,

Altinbas University

Asst. Prof.Dr. Abdullahi Abdu
IBRAHIM

School of Engineering and
Natural Sciences,

Altinbas University

Prof. Dr. Hasan Huseyin
BALIK

Faculty Of Engineering And
Architecture,

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Approval Date of Institute of Graduate Studies: ____/____/____

DEDICATION

I dedicate my dissertation work to my father and mother, may God protect them and extend their lives, to my beloved and dear wife, to my daughter and son, may God protect them for me.

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my advisor Prof. Dr. Osman Nuri Ucan for the continuous support of my Msc study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis, I thank my father and mother for their support to me, and my wife for her support and for enduring many difficulties.



ABSTRACT

EFFICIENT CONSTRUCTION COST ESTIMATION SOFTWARE USING CONVOLUTIONAL NEURAL NETWORK AND PARTICLE SWARM OPTIMIZATION ALGORITHM (PSO)

Alsabaawi, Mohanad Mohammed Hamad

M.Sc., Information Technologies, Altınbaş University,

Supervisor: Prof. Dr. Osman Nuri UCAN

Date: August, 2020

Pages: 61

Early time cost estimation plays an important role in the achievement of each construction project. Entirely parties included in the project construction; holders, workers and contributors need dependable information about rate in the primary periods of the project. There are very limited drawings and facts at this step. In this study, new method based PSO and CNN proposed for construction project cost estimation. The data first analysed by using feature extraction to reduce the size of data using PSO. The extracted features wired to the CNN that extracted high level features from input data and then wired to the regression model that used to estimate the cost of the construction project with average error rate only 0.0018. Furthermore, the proposed method compared with several studies in the same problem and the experimental results show that our method presented best results than previous studies.

Keywords: CNN, SoftMax, PSO, Construction Project Cost Estimation.

ÖZET

EVRIŞİMLİ SİNİR AĞI VE PARÇACIK SÜRÜSÜ OPTİMİZASYONU ALGORİTMASI (PSO) KULLANARAK VERİMLİ İNŞAAT MALİYETİ TAHMİN YAZILIMI

Alsabaawi, Mohanad Mohammed Hamad

Yüksek Lisans, Bilişim Teknolojileri, Altınbaş Üniversitesi,

Danışman: Prof. Dr. Osman Nuri UCAN

Tarih: Ağustos, 2020

Sayfa Sayısı: 61

Erken dönem maliyet tahmini, her bir inşaat projesinin gerçekleştirilmesinde önemli bir rol oynamaktadır. Tamamen taraflar proje inşaatına dahil edilmiştir; sahipleri, çalışanları ve katkıda bulunanlar, projenin ilk dönemlerinde oran hakkında güvenilir bilgiye ihtiyaç duyarlar. Bu adımda çok sınırlı çizimler ve gerçekler vardır. Bu çalışmada, inşaat projesi maliyet tahmini için yeni yöntem tabanlı PSO ve CNN önerilmiştir. Veriler ilk önce PSO kullanarak veri boyutunu azaltmak için özellik çıkarma kullanılarak analiz edildi. Çıkarılan özellikler, giriş verilerinden yüksek seviye özellikler çıkaran CNN'ye ve daha sonra sadece 0,0018 ortalama hata oranı ile inşaat projesinin maliyetini tahmin etmek için kullanılan regresyon modeline bağlandı. Ayrıca, önerilen yöntem aynı problemdeki çeşitli çalışmalarla karşılaştırıldığında ve deneysel sonuçlar yöntemimizin önceki çalışmalardan daha iyi sonuçlar verdiğini göstermektedir.

Anahtar Kelimeler: CNN ,SoftMax, PSO, İnşaat Projesi Maliyet Tahmini.

TABLE OF CONTENTS

	<u>Pages</u>
ABSTRACT.....	vi
ÖZET.....	vii
LIST OF TABLES	x
LIST OF FIGURES	xi
LIST OF ABBREVIATIONS	xiii
1. INTRODUCTION.....	1
1.1 CONTRIBUTIONS.....	4
2. OVERVIEW.....	5
2.1 CONSTRUCTION PROJECTS COST ESTIMATE.....	5
2.2 CONSTRUCTION PROJECTS COST ESTIMATE SOFTWARE	8
2.3 MACHINE LEARNING	8
2.3.1 Unsupervised learning	10
2.3.1.1 Clustering	12
2.3.2 Supervised learning	17
2.3.3 Semi-supervised	23
3. MATERIAL AND METHODS	27
3.1 DATASET	27
3.2 DEEP LEARNING (DL).....	27
3.2.1 Convolutional Neural Network (CNN)	28

3.2.2	Recurrent Neural Network (RNN)	32
3.2.3	Long Short Term Memory Networks (LSTM)	33
3.2.4	Particle Swarm Optimization (PSO) Algorithm	35
3.3	PROPOSED METHOD.....	37
3.4	EVALUATION PARAMETERS	39
3.5	VALIDATION USING RANDOM SUBSAMPLING.....	39
4.	EXPERIMENTS AND DISSCUSION	41
4.1	IMPLEMENTATION	41
4.2	RESULTS AND DISCUSSION.....	45
5.	CONCLUSION.....	48
	REFERENCES.....	49

LIST OF TABLES

	<u>Page</u>
Table 4.1: Our Software results.	47



LIST OF FIGURES

	<u>Page</u>
Figure 2.1: Different disciplines of knowledge and the discipline of machine learning.	8
Figure 2.2: Unsupervised learning	11
Figure 2.3: Supervised Learning.....	17
Figure 2.4: Input Features to KNN.	18
Figure 2.5: Classification using KNN.....	19
Figure 2.6: Radial Basis Function.....	20
Figure 2.9: Multi-Layer Perceptron	22
Figure 2.7: Semi-supervised learning Structure	24
Figure 2.8: Self-training Algorithm	25
Figure 2.9: Co-Learning Algorithm	26
Figure 3.1: Classification Using CNN.	28
Figure 3.2: Convolutional Neural Network [26].	29
Figure 3.3: Convolution layer.....	30
Figure 3.4: Max Pooling.....	31
Figure 3.5: average Pooling.....	31
Figure 3.6: Fully Connected layer.	32

Figure 3.7: RNN Structure.	33
Figure 3.8: LSTM based SoftMax.	34
Figure 3.9: LSTM Structure.	35
Figure 3.10: Our Cost Estimation Framework.	38
Figure 3.11: Random Subsampling.	40
Figure 4.1: Read data in MATLAB2020.	41
Figure 4.2: Costs of projects.	42
Figure 4.3: PSO Script.	42
Figure 4.4: Data processing.	43
Figure 4.5: Read Images using MATLAB2020.	43
Figure 4.6: CNN layers.	44
Figure 4.7: Example for Input Data.	45
Figure 4.8: Our Software results.	46

LIST OF ABBREVIATIONS

SVM	:	Support vector machine
KNN	:	K-nearest neighbours algorithm
NN	:	Neural Network
RBF	:	Radial Basis Function
CNN	:	Convolutional Neural Network
DT	:	Decision Tree

1. INTRODUCTION

At the current time, construction process has been improved with the effect of industrial improvements on construction area and further qualified structures are built [1]. Each construction project changes finished a sequence of phases. The start of the whole process is the preparation and design stage which creates from the initial study and is shadowed by numerous design phases. The greatest expensive phase is the application of the design with the real building [2]. Therefore, construction project managers are mainly intricate in cost control. This brands it essential for construction managers to recognize dissimilar issues that affect the correctness of project cost guesses. Numerous investigators have showed detailed studies in demand to define issues swaying the accuracy of cost guesses. Nevertheless, moderately less information on this is obtainable for the construction businesses in developing countries [3].

Machine learning (ML) is the well-known artificial intelligence field that mimics the learning techniques in human to learn from examples. ML applied to several fields such as medicine, business, and education. In previous years several studies applied machine learning and data mining techniques to estimate the construction cost estimation [14, 15].

In this section several studies that applied machine learning for construction cost estimation are presented. In [4] Yasser Elfahham presented new method for concrete structures a formula is driven for calculating the Construction Cost Index using past records. Several techniques presented for this problem such as Linear Regression, Neural Networks, and Autoregressive Time Series. In [5] Michał Juszcyk proposed study is to explain the chances and experiments of the method based on the intelligence system tools to cost estimate of construction project. The presented method is depending on the theory of nonparametric cost predication and application of ANN. In [6] Wen-der Yu et al. presented WICE is the first of-a-kind real-time theoretical cost predicating software in practice use. The validating outcomes indication that the presented software provides not only a universally nearby and punctually replying income for cost predication, but also an actual and dependable instrument for real-time decision-making

In this study, we presented new method to estimate the construction cost with accurate results and low computation time. Finally, this paper structure contains material and methods, experimental results and conclusion.

What are the building tips? It is simply the process of predicting and planning the cost of a specific construction site. An end product is a document or presentation that details the activities you typically take, the materials used, and the costs.

Project appraisal, along with labor costs and customer service, is one of the most important parts of a construction project. Job creation can be frustrating because job estimates have to be incredibly accurate. Metaphorically speaking, being too poor can be overwhelming given how your business literally makes money.

The answer to this command is simple. This is correct in the estimate so more work is required. So, Customers say so, and if you've made inaccurate or poorly constructed quotes in the past, they will tell you and in a world where it costs you extra money. By providing accurate quotes, you can be sure that both of you know what they are doing.

If you are honest with every customer, you can get more work done technically, I'm sure the customer will help you with this. If clients are concerned about the cost of this project and know if it will outweigh the overall financial impact, self-sufficiency and parrot impressions are worth it.

This gives customers a premium when they start providing the information they need for the budget. The last thing about the price is ordering the wrong material, which costs you and your customers more money.

Nobody wants to lose money ...

These two reasons are very important to get a very accurate quote. It's good for you (not hard for me) Thanks to your incredible reputation; I will save you and your customers and win new ones.

A complex system in which computers play an important role. Complex physical systems and sophisticated software systems can help people lead healthier and more comfortable lives. However, by relying on these systems, you can reduce the cost, time, and cost of these systems

in the event of a failure. Life can be wasted. Most of the complexity of these systems comes from the integration of information technology with physical and human activities. This integration considerably increases the mutual dependencies between components, people and processes and creates a complex dynamic that is not taken into account in systems of the previous generation. Engineers who understand the details of application areas and computer electronics, software, human factors, and communications should take a holistic approach to developing systems for disaster prevention.

Computer systems were introduced to the market several years ago. It has been breaking through for decades. Inquiries such as airline reservations were very common. Over time, these systems become a necessary and important part of all organizational activities. The systems engineers involved in these systems have contributed significantly to the success of this system.

Computer systems know they don't exist in a vacuum. It contains many built-in components. Usually when software and hardware are combined into a system in any form. People, software, and hardware operations interact like databases. However, its use gradually expanded over time, and computers were found to include nearly all domains and all associated applications in an organization's domain. Gradually, all the work is done directly in various applications of the computer system.

Computer Systems Engineering (ECBS) A more restrictive form of systems engineering that only deals with computer systems. It's a very important part of systems engineering, but it's still much broader than software development. Provides a complete picture of the system environment and components. Often the components of an information system are also (subordinate) information systems. ECBS focuses on examining the non-functional features of the proposed design and the need to compromise between different design options before deciding on their implementation. Compensation is seen as an important aspect of such a project as it can resolve complaints or disputes (especially commercial and technical risks).

1.1 CONTRIBUTIONS

- Developed new construction cost estimation system, this system have ability to estimate the cost of the construction project by analyzing the features about the project which contain number of features about the project.
- The new system combines PSO with spectrogram and CNN. The PSO select number of features to reduce the size of input data. This lead to remove unaffected features from input data. Then, the spectrogram of input features are plotted as figure and become input to the CNN model that's extracted high level features and classify the extracted features by using SoftMax.
- The system presented estimated cost within minimum time because its processes only important features and the redundancy features are deleted.

2. OVERVIEW

2.1 CONSTRUCTION PROJECTS COST ESTIMATE

One of the biggest challenges in the construction sector is the management of construction projects with their own characteristics and requirements. However, all project workloads, plans, budgets, etc. have something in common. Cost estimates are an important skill for engineering firms and contractors because important design decisions are based on costs.

Project owners use cost estimates to determine the scope and potential of the project and to allocate budgets. The contractor uses it to decide whether or not to bid on the project. In general, we budget with architects and engineers to ensure that projects meet scalability and financial feasibility requirements.

A good quote will help clients avoid losing money and avoiding overpayments. An important part of earned value management. Project management technology that tracks project performance based on total time and cost estimates.

It is advisable to use the offers calculated in the design phase of the project. In evaluating the project, the evaluator uses the order of the progress level, the circuit design level and the development phase of the project. The schematic design uses the schematic design to determine the feasibility of the cost estimate. In addition, technical calculations and construction documents are used in the phases of project development and construction documentation. After the 5th accreditation, there are three types of accreditation that correspond to different levels and phases of the project.

In the offer evaluation phase, we propose a proposal. Use multiple data points, e.g. B. increased design books and other direct costs. Using these data points, the reviewer determines the approximate cost of the work and provides them with other documents that are considered for work on the project. It is important for project management to keep track of the flow of information when transporting documents, to keep the power supply on the same page and to make the information available. Accurate estimates can help companies do more. When a construction company has a reputation for staying on budget, the construction company always earns more than a low-paying member who exceeds the budget.

Cost estimating this is one of the most important steps in project management. The estimate establishes the base cost of the project at various stages of its development. The cost estimate for a particular stage of project development is the forecast provided by the engineers and the estimate based on the available data. According to the American Association of Cost Engineers, value engineering is defined as an area of technical practice that uses technical judgment and experience in applying scientific principles and methods to a problem. Cost estimation, cost management and profitability.

Nearly all estimates are made using one or a combination of the following basic approaches.:

- **Production function.** In microeconomics, the relationship between the result of a process and the required resources is called the production function. Construction and production functions can be expressed by the relationship between the amount of construction and factors of production such as labor and capital. Production functions relate production or production to various labor, material, and equipment inputs. For example, the output Q is a function of the various input coefficients x_1, x_2, \dots, x_n , which can be obtained using mathematical and statistical techniques. How to find a range of input coefficient values to help minimize production costs at a given level of production. The relationship between the size of a construction project (expressed in square feet) and the amount of labor (expressed in man hours per square foot) is an example of a production function for construction.
- **Empirical cost inference.** Empirical estimation of the cost function requires statistical methods that associate the construction and operation costs of the facility with some important characteristics or attributes of the system. The role of statistical inference is to estimate the maximum value of a parameter or constant of a hypothetical cost function. This is usually done using regression analysis techniques.

- **Unit costs for bill of quantities.** The unit price is assigned to each team or activity in the quantity list. The total cost is the sum of the unit costs multiplied by the product of the quantity. The unit price method is simple in principle, but very complex to use. The first step is to break the process down into several activities. To create the installation, these tasks must be performed together. After defining these goals and estimating the quantities they represent, the total cost is found by assigning the cost to each unit and adding the cost for each activity. The level of detail of the activity history varies considerably depending on the offer.
- **Allocation of joint costs.** You can design a cost function for your job using cost allocation for existing accounts. The main idea of this method is that each cost item can be assigned a specific process characteristic. Ideally, cost sharing should be causally related to the main cost categories of the sharing process. However, in most cases, a causal relationship between allocation and cost items cannot be determined or may not exist. For example, in a construction project, core cost accounts can be classified by (1) labor, (2) materials, (3) construction equipment, and (4) construction supervision. Construction (5) General office expenses. These baseline costs can be prorated across the fragmented project activities.

2.2 CONSTRUCTION PROJECTS COST ESTIMATE SOFTWARE

Predication of cost of construction projects is an actual complex procedure covering numerous variable issues. This ability is not simply learned. Proper study, training and knowledge are wanted to develop capable in building project cost predication. There are numerous groups that can have important influences on project costs. The estimator should be conscious of them and correctly assess their belongings, previous to confirming the cost predicate.

Cost predicates are also beneficial when analyzing actions that increase any feature of a construction's performance. These processes upsurge the project budget, but they can produce better welfares over time. In the last years, several automatic systems based on machine learning and data mining techniques are presented for construction projects cost estimation [7].

2.3 MACHINELEARNING

Machine learning is a branch of artificial intelligence that aims to do this so that machines can do their work with skill intelligent program. Statistical forms of learning the backbone of intelligent software used for development machine intelligence. Because machine learning algorithms requires data to learn, discipline must be in contact database discipline. There are also known terms such as data discovery knowledge (KDD), data mining, pattern recognition. One wonders what the oldest looks like the image in which this connection is shown.

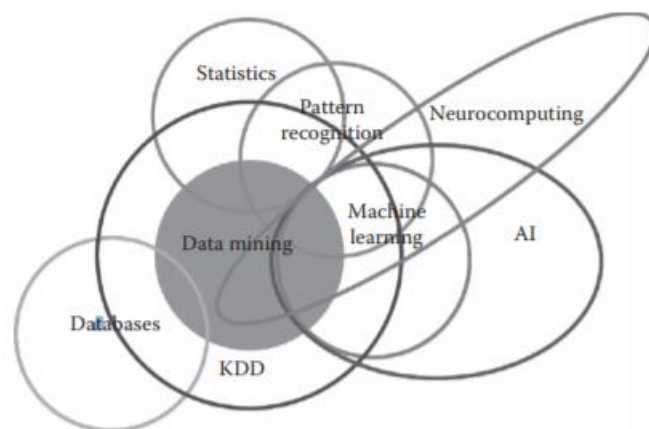


Figure 2.1: Different disciplines of knowledge and the discipline of machine learning.

When you tag a face on a Facebook photo, the AI works in the background and identifies the faces in the picture. Face recognition has become popular in many applications. View images with human faces. Why only human faces? There are many uses for learning things like cats, dogs, bottles, cars, etc. We have independent cars that roam our streets to learn things in real time to control the car. When traveling, they use Google's instructions to determine traffic conditions in real time and to follow the best route currently suggested by Google. This is another application of real-time object recognition technology. Let's look at an example of the Google translation app that we usually use when we go abroad. With Google's online translation app on your phone, you can reach locals and speak a foreign language. There are many applications of artificial intelligence that we use in practice today. In fact, any artificial intelligence has been used in many parts of our lives, even without our knowledge. Amnesty International can now perform very complex tasks with great precision and speed. Let's discuss an example of a complex task to understand what skills an AI application will develop for your customers today. The development of artificial intelligence applications began today using the ancient tradition of statistical techniques. Schools must have used linear interpolation for the forecast Future value. There are many other statistical techniques that have been successful Applied in the development of so-called programs for artificial intelligence [29, 30]. We say "what is called" because the software for artificial intelligence Today we are more sophisticated and use technologies that go far beyond statistics Technologies used in early AI programs. Some examples of statistical techniques for developing artificial intelligence applications in those days and still in practice are listed here:

- Regression
- classification
- Clustering
- Probability theories
- Rees decision trees

We've just scheduled a few basic techniques here to help you get started with AI without frightening you with the breadth that AI needs. Use these statistical methods when developing artificial intelligence applications based on limited data. Today, however, the data are plentiful. Statistical techniques are not very useful for analyzing the type of big data, as they have certain inherent limitations. Therefore, additional progressive approaches like deep learning have been industrialized to explain several of these composite problems. If we continue with this tutorial, lead to apprehend what machine learning is and how it is applied to problems or to improve complex AI softwares.

2.3.1 Unsupervised Learning

Let us consider a machine that receives a certain order of inputs $x_1, x_2, x_3, \dots, x_t$ where x_t is the sensual contribution to time t . This input, the functionality of which we will frequently call, could resemble an image on the retina, the pixels of a camera, or a sound waveform. It could also resemble to less obvious carnal data, such as the words in a news article or the list of articles in a superstore basket.

There are four dissimilar types of machine learning. With supervised learning, machine1 also receives an arrangement of wanted outputs y_1, y_2, \dots , and the purpose of the machine is to learn how to yield the precise output on a new input. This output can be a class name (in classification) or a actual amount (in regression).

Execution of activities a_1, a_2, \dots These activities touch the national of the situation, which in try principals to scalar plunders r_1, r_2, \dots The aim of the machine is to acquire to action in like a method that the future plunders it obtains (or minimizes the penalties) during its life are maximized. Strengthening learning is carefully connected to the parts of choice model and control theory. The important problems inspected in these parts are repeatedly officially equal and the answers are the same, though dissimilar features of the problem and the solution in general are emphasized.

The third type of machine learning is carefully correlated to game theory and make a sweeping statement the improvement of learning. Here, too, the machine obtains the activities created and obtains plunders. The situation in which the machine interrelates is not a fixed domain, but can

cover additional machines that it can also feel and action, obtain plunders and learn from. Thus, the purpose of the machine works to increase the rewards by looking at the future actions of current and current devices. Although game theory costs a lot of work in systems for simplicity, the dynamic state with many fitting machineries leftovers active in a difficult search area [31, 32].

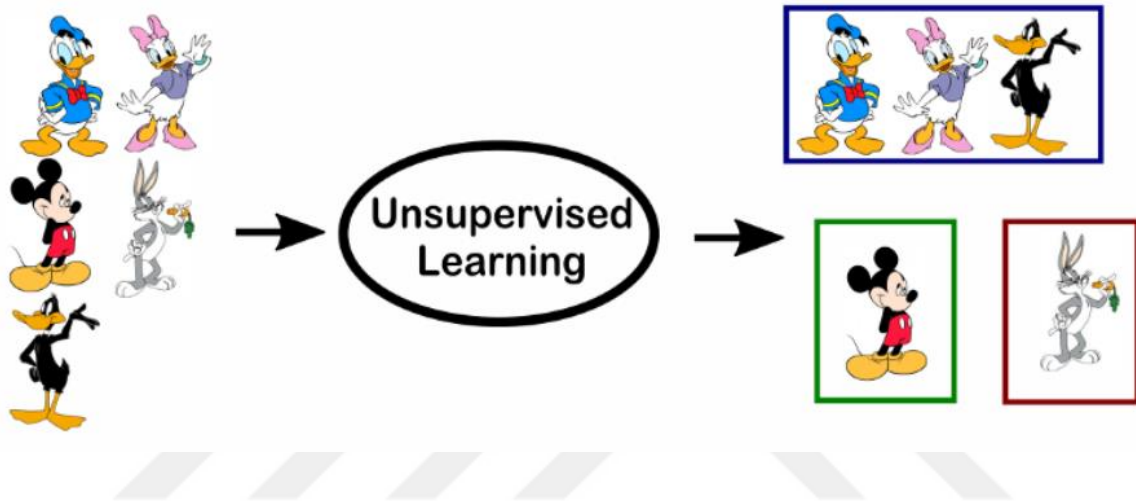


Figure 2.2: Unsupervised learning

As shown in figure 2.2, unsupervised machine learning purports to discover before nameless designs in data, but greatest of the period these decorations are deprived estimates of what supervised machine learning can attain. Moreover, meanwhile you don't know what the results must be, there is no way to control how precise they are, creation supervised machine learning additional applicable to real-world problems.

The superlative time to apply unsupervised machine learning is when you don't have data on target outcomes, like decisive a target marketplace for a completely new creation that your commercial has not ever sold beforehand. Though, if you are annoying to become a healthier sympathetic of your current customer base, supervised learning is the best method.

2.3.1.1 Clustering

- **K-means**

K-mean process is an iterative procedure that stabs to divider the dataset into K pre-defined separate non-overlapping subcategories where every feature fact belongs to only one set. It attempts to brand the inter-cluster feature facts as related as likely while also care the clusters as dissimilar as conceivable. It allocates feature facts to a cluster like that the amount of the squared distance between the facts opinions and the cluster's centroid is at the smallest. The fewer difference they have inside clusters, the extra similar the feature facts are inside the same cluster. [33, 34]. The way k-means algorithm works is as follows: 1) Identify amount of clusters K. 2) Modify centroids by first scuffling the database and then arbitrarily extracting K feature facts for the centroids deprived of standby. 3) Save repeating pending there is no alteration to the centroids. i.e task of data opinions to clusters isn't altering.

- **Principle Compont Analysis (PCA)**

PCA is basically fair a organize alteration. The innovative feature are plotted on an X-axis and a Y-axis. For two-dimensional feature, PCA pursues to alternate these two battleaxes so that the novel axis X' lies lengthways the way of supreme difference in the features. PCA needs that the axes be vertical, so in two sizes the choice of X' will control Y'. You get the transformed feature by interpretation the x and y standards rotten this new group of axes, X' and Y'. For additional than two dimensions, the initial axis is in the way of greatest difference; another, in way of the next-most difference; and so on [35].

- **Density-based Optimal Clustering (DOC):** DOC projective clustering is considered a hybrid approach, which combines both bottom-up and top-down approaches. Essentially, DOC considers a projective cluster, and it defines an optimal projective cluster in a mathematical way for this purpose. The projective cluster in DOC is established by determining a strong clustering tendency of a subset of instances C towards a subset of intervals D. In DOC, this is formed as a pair (C, D), and in order to create the aimed clustering process, DOC must identify the optimal pairs (C, D). Here, identifying the

optimal pairs is done by generating a small subset S , by any valid means, for example, random sampling. Instances in both C and S are compared and the absolute result of their difference should be less than or equal of the fixed-length, ω , which is a fixed subspace length provided by the user. This procedure is repetitively done, and both C and S are continuously obtained, using random sampling or any alternative method, until the optimal result is achieved. The minimum number of instances which can create a class, α , is also needed for the operational process of DOC. The minimum allowed density for a cluster is then determined by both parameters α and ω . It is also suggested that a trade-off should be made between the number of intervals and subspaces. For this specific purpose another parameter is used, which is usually referred to by β . This balancing process is at extreme importance for DOC, as the time complexity grows exponentially with respect to the intervals number and linearly with the cells number [18].

- **MAFIA:** Like ENCLUS, MAFIA is another variant of CLIQUE. However, unlike CLIQUE, MAFIA improves the performance of the outcomes by deploying an adaptive grid size. This grid size is obtained from the data distribution of a given dataset. The general steps used by bottom-up methods are also used by MAFIA where initially the whole space is divided into a set of subspaces. The minimum required number of intervals in MAFIA is determined by deploying a histogram. The same idea used in CLIQUE of merging adjacent dense cells into larger regions, which form a cluster, later on, is also used in MAFIA. Nevertheless, space partitioning is done based on data distribution in this case, which forms the adaptive nature of the algorithm and allows more accurate results to be produced. MAFIA requires thresholds for both area density and adjacency merging. These thresholds will allow merging adjacent regions within a specified threshold value; these merged regions form the clusters later on. It is important to mention that MAFIA is sensitive to these threshold values. In the usual operational process, MAFIA deploys a default grid as seed, where the adaptive grid is then formed based on it. Like CLIQUE and ENCLUS, MAFIA can operate based on different input scenarios, such as large datasets, different sizes, and shapes of subspaces, and it can also produce overlapping clusters as well. However, unlike

CLIQUE and ENCLUS, MAFIA outperforms these two algorithms in terms of time complexity, yet, it is crucial to mention that the time complexity of MAFIA also exponentially grows when the number of intervals increases [19].

- FINDIT: This algorithm refers to Fast and Intelligent Subspace Clustering Algorithm using Dimension Voting. FINDIT is a top-down approach that deploys the essence of voting in its operation. The concept here relies on a certain distance measure referred to by Feature-Oriented Distance (FOD). FINDIT identifies instances that agree on features with a specified threshold value, ϵ . In high-dimensional data, instances may get a close match with several features but not all of them. The initial step in FINDIT is achieved by choosing two small sets, randomly created, to be clusters' medoids; this stage is called the sampling stage. In the next stage, FINDIT uses the generated medoids for calculating FOD to determine the correlation, which allows the creation of clusters. The threshold is then iteratively increased until the final clustering shape comes to a stable state. This stage is called the cluster-forming stage. Finally, FINDIT allocates the dataset instances to the formed medoids considering all identified subspaces. Generally, two main parameters are used by FINDIT, namely the minimum permissible instances number per cluster, and the minimum permissible distance between any two clusters. One of the main disadvantages of FINDIT is the iterative nature of finding the optimal, which increases the time complexity of the algorithm. It is possible in many applications which use FINDIT to find a proper sampling technique integrated with the algorithm to improve its usability and running time quality [20].
- ORCLUS: This algorithm uses the correlations contained or generated as an inter-attribute in a given dataset. Initially, ORCLUS outlines clusters arbitrary then it assigns cells to the closest cluster in an iterative way; this stage is referred to as clustering assignment. The next stage followed by ORCLUS is achieved by identifying subspaces associated with the generated clusters. This is mathematically obtained by choosing the eigenvectors with the smallest eigenvalues after calculating the covariance matrix for each cluster. This produces some clusters that are close to each other and can be merged together. The mathematical computation of ORCLUS creates a high complexity, which makes the algorithm deploy some sort of sampling scenario to enhance the performance. Nevertheless, possible missing small clusters are inevitable in this case [21].

- **PROCLUS:** Projected Clustering (PROCLUS) algorithm was the oldest top-down approach. PROCLUS uses similar principles of those deployed by ORCLUS. The main idea of dataset sampling, setting up a number of mediods, and then iteratively improving the clustering process is also applied in PROCLUS. A suitable greedy algorithm is used for choosing a number of possible mediods. These mediods should be relatively not close to each other. During this stage, PROCLUS makes sure that, at least, one class is representing each cluster. Generally, this phase is referred to by clustering initialization. Once the initialization step has been concluded, PROCLUS deploys a number of randomly selected mediods, k , and iterate over the reduced dataset to evaluate whether the clustering process is improved or not. A specific parameter is used to maintain the quality of the clustering process, namely the average subspace dimensionality, l . Each medoid has to associate $k - l$ features. Identifying the targeted subspaces for each medoid is followed by assigning data points to these medoids; this is done using Manhattan segmental distance. The process is concluded by eliminating any medoid with a minimal number of data points. This stage is referred to by the iteration phase. The final stage in PROCLUS is achieved by refining the created clusters where the assignment of data points may be shuffled to come up with better quality clusters. This stage is called the refinement phase. PROCLUS requires working with equally divided subspaces and with the present average number of features. Overlapping in PROCLUS is a case to consider, as each set of data points are ultimately associated with one cluster. The running time of PROCLUS is usually little faster than CLIQUE and its variant ENCLUS considering large datasets. Similar to iterative approaches, PROCLUS is sensitive to the input parameters, which affect the quality of the clustering process. In addition, medoids elimination may completely discard some useful clusters in the process [22].
- **δ -Clusters:** This algorithm deploys the coherence between instances subset and attributes subset. The idea is based on calculating this coherence and using it as a distance measure for clustering purpose. The relationship that can be observed with coherent data samples is used for generating a proper cluster, as it is possible to discover instance based on their coherence. It is possible to use any valid coherence measure; however, some researchers have used PearsonR for such purpose [50]. Using δ -clusters, cluster seeds are initially created, then the algorithm attempts iteratively to improve these clusters using a random

shuffling technique, which tries different data samples and attributes to get a better clustering quality. This process stops when the iterative process produces no further improvements. Two main parameters are needed for the operational process of δ -clusters algorithms, namely the size of the cluster and the total number of considered clusters. It is important to mention that the time complexity of δ -clusters depends on these two parameters. In addition, while implementing the δ -clusters method, it is recommended to maintain pre-knowledge or some sort of speculation of cluster size, as it makes difference with large datasets. Unlike other top-down approaches, δ -clusters deploys the coherence as a measure, which makes it a potential method for a number of applications, especially in the field of medicine [23].

Despite the diversity of clustering approaches, stumbling with unsatisfied clustering outcomes is a common problem, especially when very dense or sparse datasets are considered. Nowadays, with many different sophisticated systems being in use, the complexity of mined datasets has become generally very high; thus, avoiding the use of dimensional data makes current systems very impractical. Meaning that any enhancement solution should consider the use of high dimensional datasets, as they are the commonly produced output in many modern systems. In this regard, the advancement in computerized processes has enabled the deployment of soft methods that relatively have a lower implementation cost as well a more efficient clustering outcome. Mainly, for this reason, researchers have put forward different soft clustering enhancement methods, especially in the past two decades. Soft subspace clustering is one of the recent trends deployed as a solution for the persistent high dimensionality problem. The method itself may use different algorithms aiming at reducing the dimensionality and achieving the desired outcomes.

2.3.2 Supervised learning

Supervised learning is a machine learning model for obtaining information about input-output relationships from a system depend on a specific set of corresponding input-output learning models. Since the output is considered a label for input or monitoring data, an input-output training sample is also called tag data or monitored data. It is also sometimes called teacher learning, learning from tagged data, or inductive machine learning. The aim of supervised learning is to form an artificial model capable of learning the correspondence between inputs and outputs and to forecast the yield of the model with new inputs. If the output accepts a finite set of discrete values indicating the class names of the input, the learned mapping leads to the classification of the input data. If the output takes continuous values, this leads to a regression of the input. Information on input-output relationships is often represented by the parameters of the learning model. If these parameters are not directly available in the training models, an apprenticeship system must go through an estimation process to obtain these parameters. Different forms of unsupervised learning, supervised learning training data require monitored or informed information, while unsupervised learning training data is not supervised because it is not labelled (this is the case) (simply inputs). If an algorithm applied supervised and neglected training data, it is called a semi-supervised learning algorithm. If an algorithm actively requires a user / teacher for labels in the training process, supervised iterative learning is named active learning [36, 37].

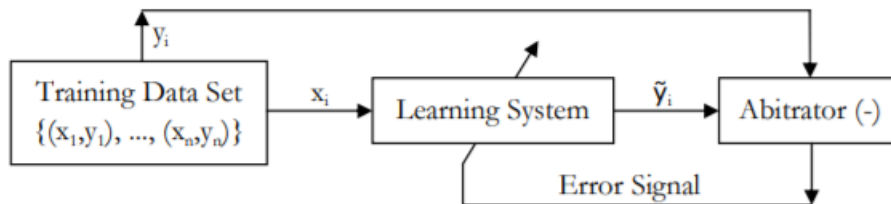


Figure 2.3: Supervised Learning.

Figure 2.3 shows a graph showing supervised learning. In this diagram, (x_i, y_i) is a monitored learning sample, where "x" represents system inputs, "y" represents system outputs (i.e. monitor

or tag the input x) and "i" is the index from the training sample. During the monitored learning process, the input of learning x_i is sent to the learning system and generates the resulting learning system \tilde{y}_i . The learning system outputs of the learning system are then compared to the basic fact mark y_i by a judgment that calculates the difference between them. The difference, which is called the error signal in this diagram, is sent to the learning system to adjust the learner's parameters. The goal of this learning process is to obtain a set of optimal parameters for the learning system that can reduce the differences between \tilde{y}_i and y_i for all i . H. Reducing the total error in the entire training record.

- **K-Nearest Neighbors (KNN)**

KNN is a humble technique that supplies all obtainable bags and classifies new suitcases depended on a likeness degree (e.g., distance functions). KNN has been applied in statistical predication and pattern classification previously in the start of 1970's as a non-parametric method [38].

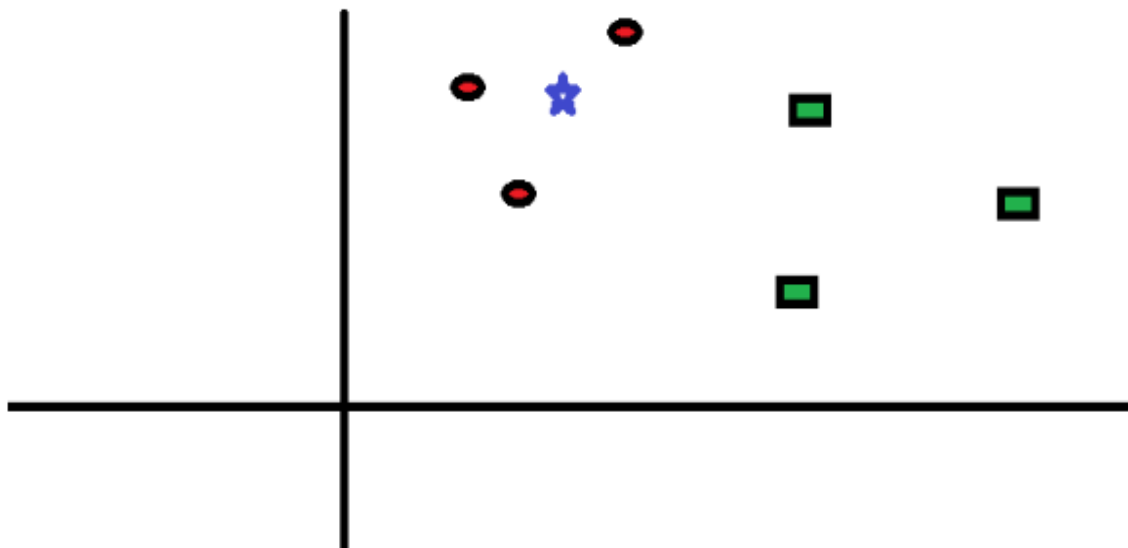


Figure 2.4: Input Features to KNN.

KNN achieved the distance formula to calculate the distance between every feature point and the validation feature. It then discovers the likelihood of these facts existence alike to the validation features and classifies it depended on which facts part the maximum probabilities.

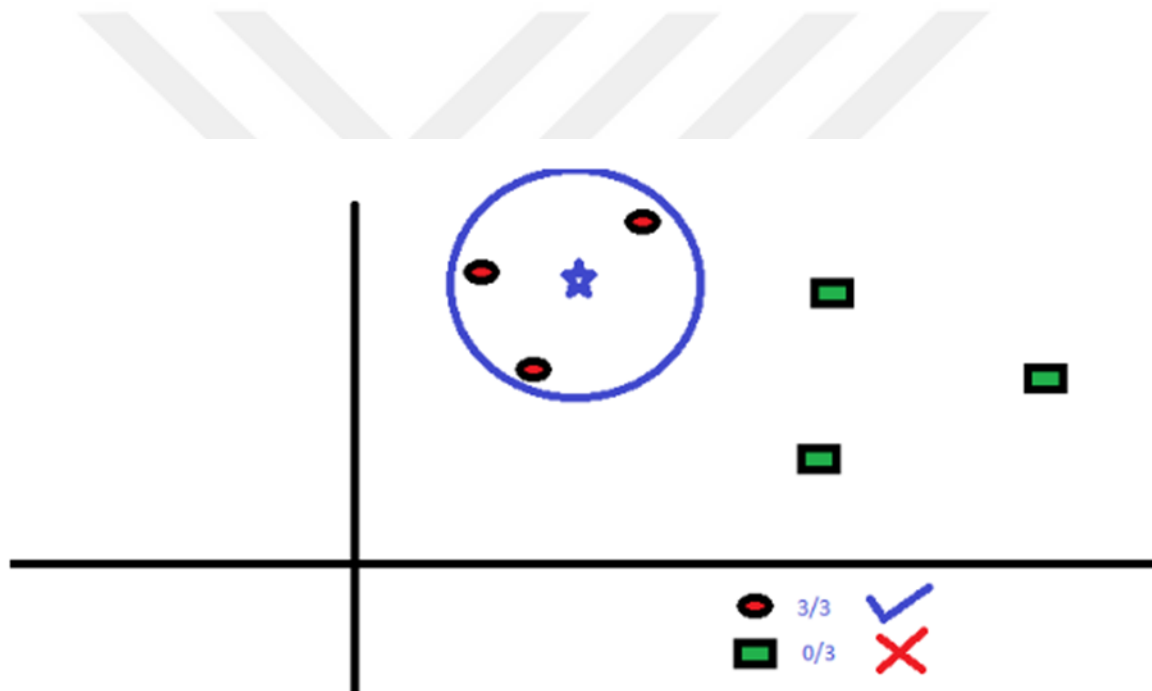


Figure 2.5: Classification using KNN.

- **Radial Basis Function (RBF)**

A RBFN is a specific kind of NN. In this thesis, it will be explain in gistapplies as a non-linear classifier.

Commonly, when person's conversation around or "ANN" they are denoting to the MLP. Individually neuron in an MLP incomes the weighted amount of its input standards. That is, apiece input cost is increased by a constant, and the outcomes are wholly summed collected. A

lone MLP neuron is a humble linear classifier, but multifaceted nonlinear classifiers can be constructed by mixing these neurons into a net.

To me, the RBFN method is additional instinctive than the MLP. An RBFN achieves classification by calculating the input’s resemblance to instances from the training group. Separately RBFN neuron supplies a “prototype”, which is fair one of the instances from the training group. When we poverty for classifying a new input, individually neuron calculates the Euclidean distance among the input and its example. Unevenly language, if the input extra carefully looks like the lesson A examples than the lesson B examples, it is identified as lesson A [39, 40]. The construction of RBF shown in Figure 2.6.

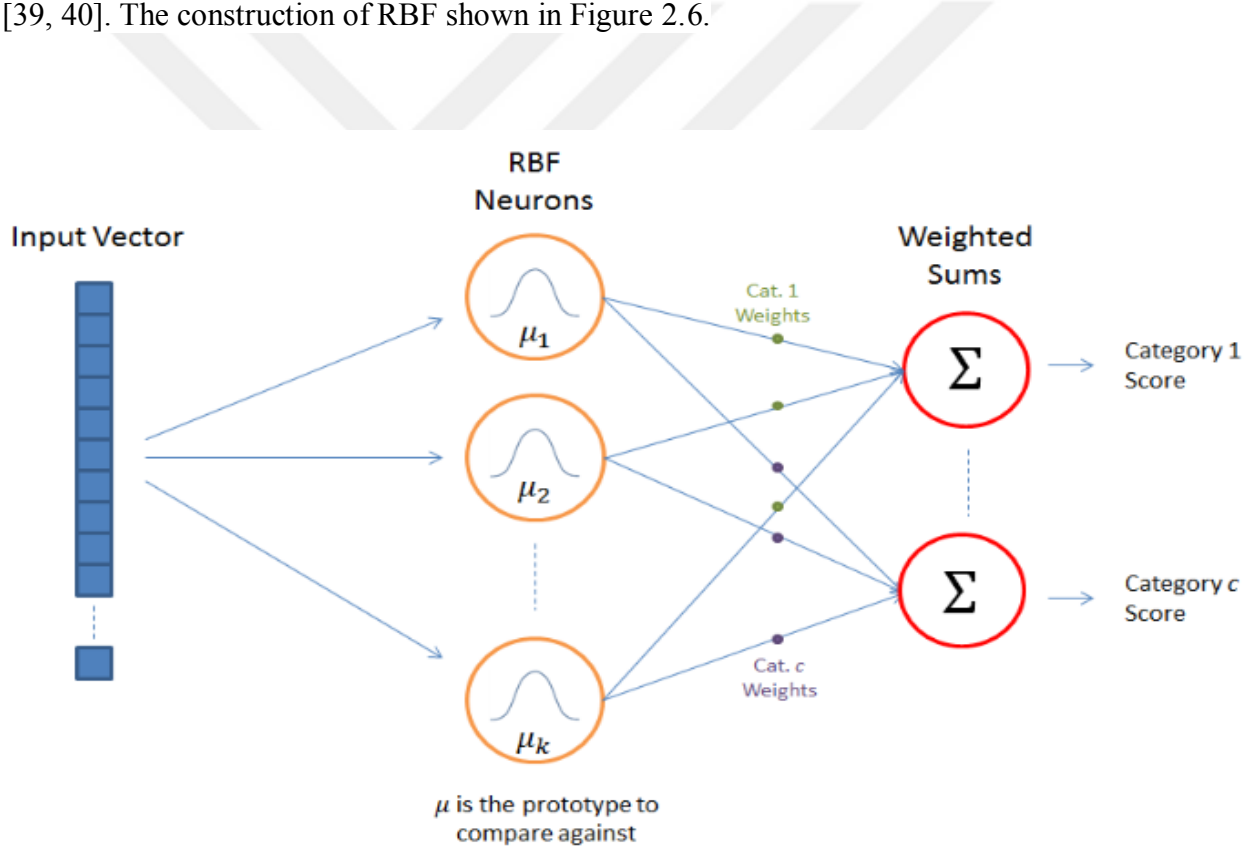


Figure 2.6: Radial Basis Function.

The beyond diagram displays the characteristic building of an RBF Network. It contains of an input array, a layer of RBF neurons, and an output layer with one nodule per group or lesson of data.

The input features are the n-dimensional features that you are annoying to categorize. The whole input feature is exposed to each of the RBF neurons see equation (2.1).

$$\phi(r) = \exp\left(-\frac{r^2}{2\sigma^2}\right) \quad (2.1)$$

Where r is a receptor and σ value of shaping parameter.

$h(x)$ is the Gaussian system with the variable r (radius) and c (center) distinct distinctly at apiece RBF component. The training procedure is depending on regulating the limits of the net to repeat a traditional of input-output designs. There are 3 kinds of variables; the weight w among the concealed bulges and the output bulges, the center c of individually neuron of the hidden layer and the component width r . The manufacture of the net covers of a collection of nodes, unique each collection that we are maddening for categorizing. Unconnectedly produce node computes a kind of entire for the connected collection. Characteristically, a group excellent is comprehensive by broadcast the input to the collection with the highest notch.

- **Decision Tree (DT)**

A tree has numerous similarities in actual life, and goes out that it has prejudiced an extensive share of ML, cover both regression and classification. In choice examination, a DT can be applied to visually and openly signify choices and decision making. As the designation energies, it usages a tree-like model of decisions [41]. The DT Algorithm Pseudocode listed below:

- Home the best element of the database in the root of the tree.
- Splitting the training group into subsections. Subsections should be complete in like a method that every subsection covers feature with the similar worth for a feature.
- Recurrence stages 1 and step 2 on every subsection till catch foliage bulges in all the twigs of the tree.

- **Multi-Layer Perceptron**

Literature anatomy reveals a persistent implementation of Feed Forward NN, from among the several categories of link for ‘artificial neurons’. A type of feed forward NN technicality is the MLPNN. The design of MLPNN is shown in Figure 2.8.

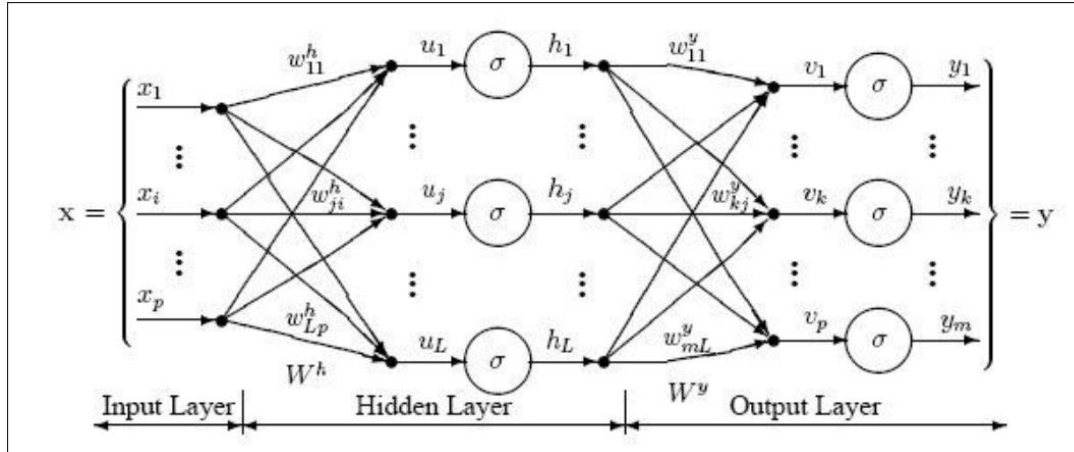


Figure 2.7: Multi-Layer Perceptron

One of the most famous models in ANN is MLPNN.

Three Types of layers in the ‘MLPNN’: the first type is an input layer, the second one is a hidden layers. The third one is output layer type (several nodes in each layer). Information passes from one node to the next node. Outer nodes transform signals to the input layer. The ‘output of first layer’ passes to second layer over “weighted connection line”. It accomplishes computations and sends the result to final layer (output) over weighted links. After that the output of second layer is send to last layer, for a final result [30].

The working steps of (MLPNN) are given in the following:

- (1) Input data is given to the first layer for treatment which leads a divine output.
- (2) The error is calculated by the difference between the predicted output & actual output.
- (3) Back-Propagation algorithm has used to adapt the weights.
- (4) For adjusting weights, it begins from weights between last layer nodes and last hidden layer nodes and works backwards through the network.
- (5) When BP is finished, the operation starts again.
- (6) The operation is renewed until the error between actual & predicted output to reach the least difference.

2.3.3 Semi-supervised

Every machine learning algorithm needs data to learn. But even with so much data around the world, including text, images, time series and more, only a small part is really categorized, algorithmically or manually.

Most of the time, we need disaggregated data to perform supervised machine learning. In particular, we use it to predict the display of each form data point. Since the data tells us what the classification should look like, we can calculate the difference between the forecast and the rating and then reduce that difference.

As you know, there is another class of algorithms called uncensored algorithms that do not require classifications, but can learn from unrated data. Uncensored learning is often a good way to discover new patterns in the data set and to group data into multiple categories based on many benefits. Common examples of K-means and latent dirichlet allocation (LDA) algorithms are.

Now imagine that you want to create a template for classifying text documents, but give the algorithm a hint to create categories. You only want to use a very small part of the text documents with the label because not every document is classified. At the same time, you want your model to classify unclassified documents as accurately as possible based on the existing documents.

Fortunately, there is a class of algorithms called semi-supervised algorithms that can learn from partially categorized data sets.

Text documents are a particularly common form of unclassified data, including files and amounts of text, programs, books, blogs, etc. that are often unclassified. We know that naming a lot of data is very expensive [42, 43]. (For example, imagine how long it would take someone to read ten thousand articles to assign a single specific category to this text document.)

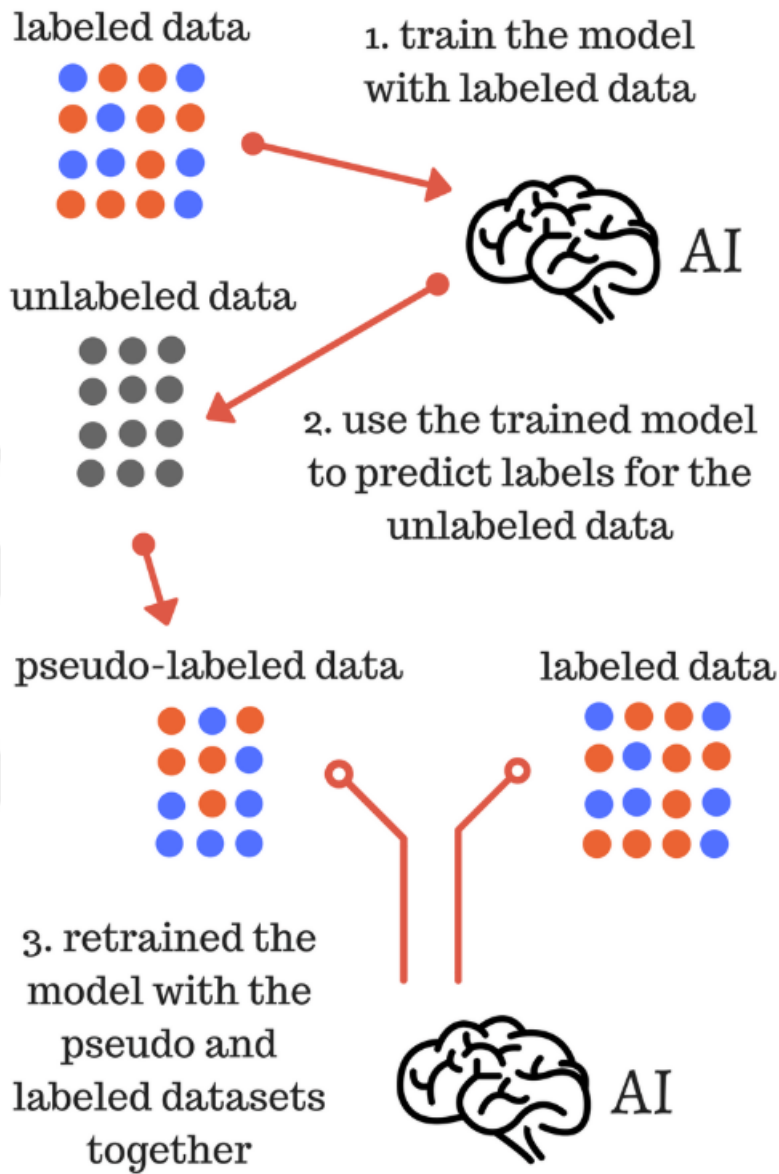


Figure 2.8: Semi-supervised learning Structure

- **Self-training**

A self-determined approach that can be applied to various scenarios where gold posters are unavailable (or partially available) and to human efforts to add costly annotations to the posters. The goal is to recover useful information from the large set of unclassified data to improve exercise performance. In the traditional self-training framework (algorithm 1), the mark-up tool is first configured with a group Gold cases. Then this day Used to tag and tag a group of unclassified data Trust is assessed for each unnamed entity. Instances that are automatically named by the training group is given more confidence. The brand will be recycled in the updated training a data set is used to mark and select unclassified elements Examples of the remaining record. This process is repeated up to a certain number of cases it is selected and added to the training record. How to measure trust it is crucial for a self-training approach. Traditionally A self-coaching approach generally requires human effort to develop specific measures based on trust on inference and application scenarios. This, However, it is difficult and effective the solution can depend on many factors However, it has not been studied well [44, 45].

```

Input: dataset train  $T$ , dev  $D$ , budget;
         unlabeled data  $U$ 
randomly shuffle  $U$ ;  $i \leftarrow 0$ ;
while  $i < budget$  and  $U$  is not empty do
     $tagger \leftarrow train(T, D)$ ;
     $tag(tagger, U)$  ;
     $x \leftarrow select(U) | \max_x confidence(y)$ ;
     $T \leftarrow T \cup (x, y)$ ;
     $U \leftarrow U \setminus (x, y)$ ;
     $i \leftarrow i + 1$ ;
end
Output:  $tagger$ ;

```

Figure 2.9: Self-training Algorithm

- **Co-Learning**

The traditional method of co-training create classifiers for different views and exchange predictions of unlabeled data with great confidence to extend data training in two views at each training session. After that, several other developments were developed. This Co-training variants can be roughly divided into two categories paradigms. One paradigm is to follow the iterative co-training process, but to label the unlabeled samples using different methods with certain confidence criteria each iteration. The other is to integrate additional information from other angles as a concept of regularization of learning Target and transforms the semi-monitored Multiview problem into a new optimization problem [46, 47].

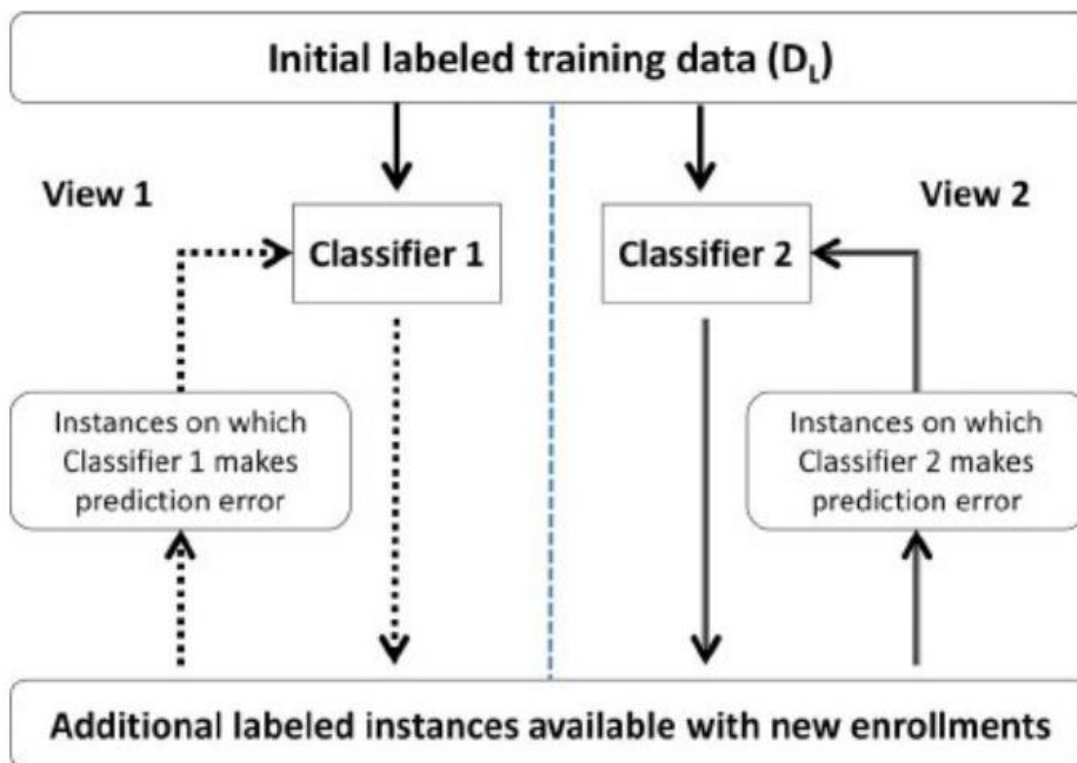


Figure 2.10: Co-Learning Algorithm

3. MATERIAL AND METHODS

3.1 DATASET

The dataset that used in this study consist from 169 instances that collected between 2009-2012 from Gaza Strip Gaza Strip building projects. The dataset collected by Omar M. Shehatto and represented all the features that affected the construction cost. The dataset consist from 11 features: Area of typical floor, Use of building, number of Floors, Footing Type, Slab Type, number of Elevator, Ex. Finishing, Air-conditioning, Tilling type, Electrical type, and Mechanical type. Finally, the response of these features represented the cost of the project in USD dollar.

3.2 DEEPLARNING(DL)

DL is a subfield of ML worried with procedures enthused by the construction and purpose of the brain named ANN. The experts in the arena need thoughts of what DL is and these exact and nuanced viewpoints hut a huge of bright on what DL is all around [55].

One of the greatest mutual AI methods applied for dispensation big data is MLg, a self-adaptive technique that becomes progressively improved examination and designs with knowledge or with anew extra data. If a numerical expenditures business required detecting the incidence or possible for deception in its scheme, it might employment ML tools for this function. The computational technique constructed hooked on a computer perfect will procedure all dealings trendy on the numerical stage, find designs in the facts set and opinion out any irregularity noticed by the design [56].

DL, a subsection of ML, uses a hierarchical level of ANN to transmit out the procedure of ML. The ANN is built similar the human brain, with neuron bulges linked together like a web. Though classical packages shape examination with information in a lined way, the hierarchical purpose of DL schemes allows machineries to procedure data with a nonlinear method [40]. With the arrangement of sufficient such alterations, very multifaceted jobs can be learned. For classification tasks, advanced layers of demonstration intensify features of the input that are significant for taste and overpower immaterial variations new learning algorithms and buildings that are presently existence industrialized for deep neural networks will only hurry this growth.

Deep learning contains number of techniques such as: deep auto-encoder, convolutional neural network and recurrent neural network [57].

3.2.1 Convolutional Neural Network (CNN)

Convolutional layers are the main structure chunks applied in CNN. A convolution is the humble presentation of a filter to an input that outcomes in a function. Recurrent application of the same filter to input outcomes in a diagram of activations named a data map, signifying the sites and forte of a noticed eye in a contribution, like an image.

The novelty of CNN is the capability to mechanically learn a great amount of filters in parallel specific to a training data below the restraints of an exact prognostic demonstrating problem, like as image recognition. The outcome is extremely exact data that can be identified wherever on input picture [24].

Like to how a child learns to recognize items, it require to expression an technique millions of images previously it is be capable to simplify the contribution and brand forecasts for picture it has not ever seen beforehand.

Machines ‘see’ in a dissimilar way than we do. Their world contains of lone facts. Each picture can be signified as 2-dimensional vector of numbers, recognized as pixels.

Then the detail that they observe pictures in a dissimilar method, doesn’t nasty we can’t sleeper them to identify forms see figure 3.1.

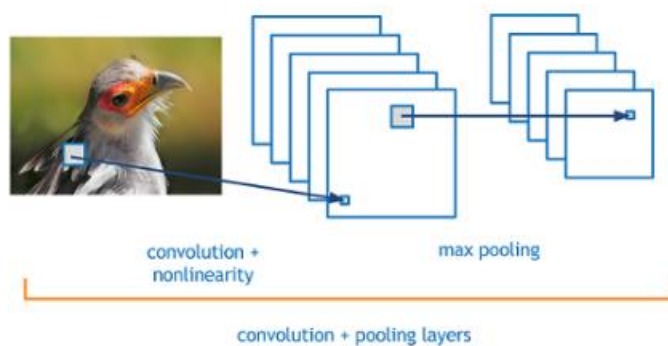


Figure 3.1: Classification Using CNN.

On the other hand, the structure of CNN presented in the figure 3.2

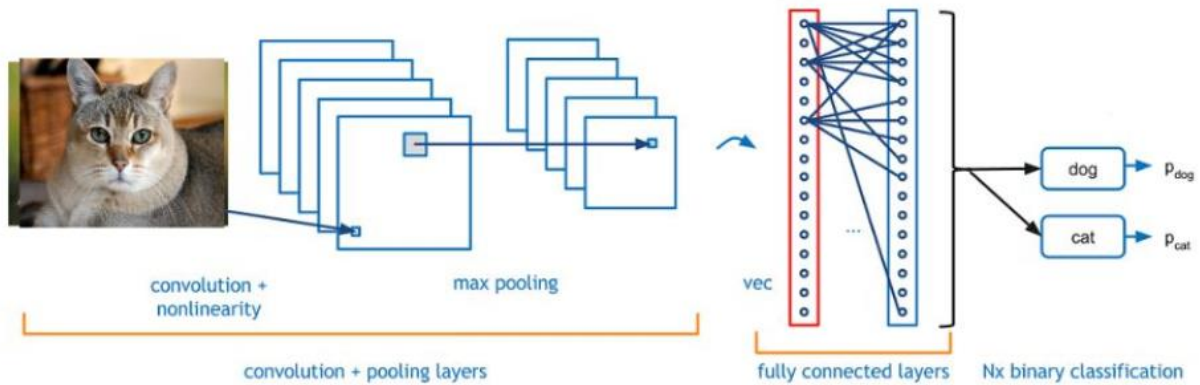


Figure 3.2: Convolutional Neural Network [26].

- **Convolutional Layer**

Afterward the input layer established the input pictures, these pictures will be convolved by the convolution layer, which contains of a group of dissimilar learnable filter banks of dimension $m \times n \times r$. every filter is smaller than the input data, but it spreads finished the full deepness of the input data size. Throughout the convolution, every filter bank will be slipped crossways the tallness and breadth of the input pictures, making a 2D data map of that filter. When we transparency the filter crossways the input features see Figure 3.3.

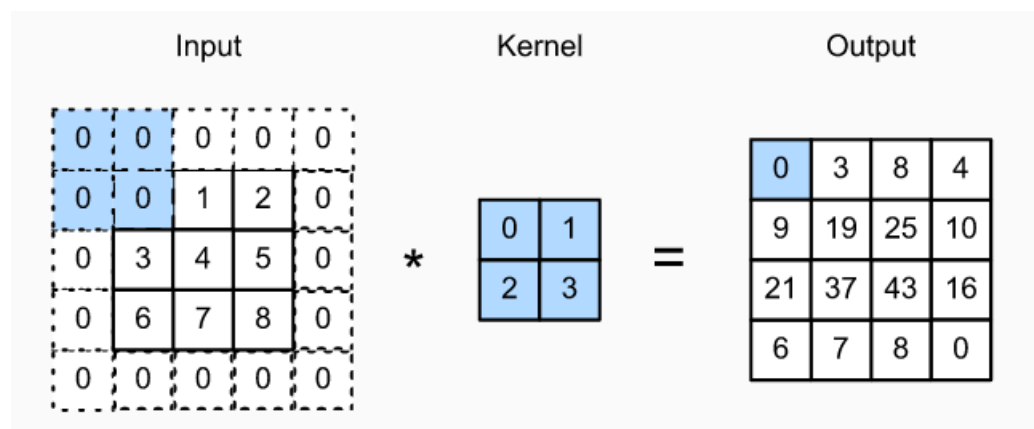


Figure 3.3: Convolution layer.

- **Pooling**

The pooling layer works upon each feature map individually to generate a different group of the similar amount of pooled feature maps.

Pooling includes choosing a pooling process, abundant like a filter to be practical to feature maps. The dimension of the pooling process or filter is minor than the scope of the mouth map; exactly, it is nearly continuously 2×2 pixels practical with a pace of 2 pixels.

This income that the pooling layer will continuously decrease the scope of each feature map by a issue of 2, e.g. respectively size is halved, decreasing the amount of pixels or standards in every feature chart to one neighborhood the scope. For instance, a pooling layer practical to a feature map of 6×6 (36 pixels) will consequence in an production pooled feature chart of 3×3 (9 pixels).

The pooling process is stated, somewhat than erudite. Two mutual purposes applied in the pooling process are:

- Average Pooling: Compute the average cost for a piece cover on the feature map.
- Maximum Pooling: Compute the maximum cost for individually cover of the feature map.

The outcome of consuming a pooling layer and making depressed tested or pooled feature maps is an abridged form of the features noticed in the input. They are beneficial as minor vicissitudes in the

position of the feature in the input noticed by the convolutional layer will consequence in a pooled data chart with the data in the similar position. This competence additional by pooling is named the replica's invariance to native conversion see Figure 3.4.

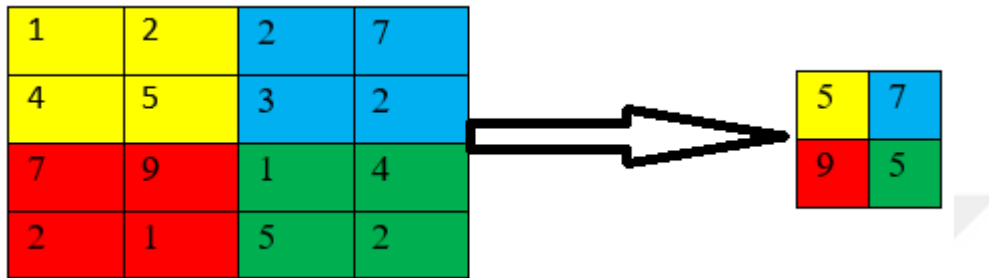


Figure 3.4: Max Pooling

Moreover, the alike instance practical with regular pooling and available additional consequences see Figure 3.5.

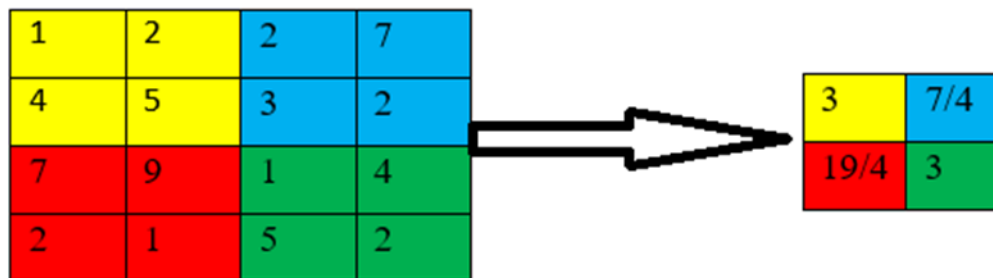


Figure 3.5: average Pooling.

- **Fully connected layer (FC)**

FC in a NN is those layers anywhere all the contributions after one coating are linked to each activation unit of the following layer. In greatest general ML replicas, the previous scarce layers are FC layers which accumulate the feature removed by earlier layers to procedure the last output. It is the additional greatest time overwhelming layer next to Convolution Layer sees Figure 3.6.

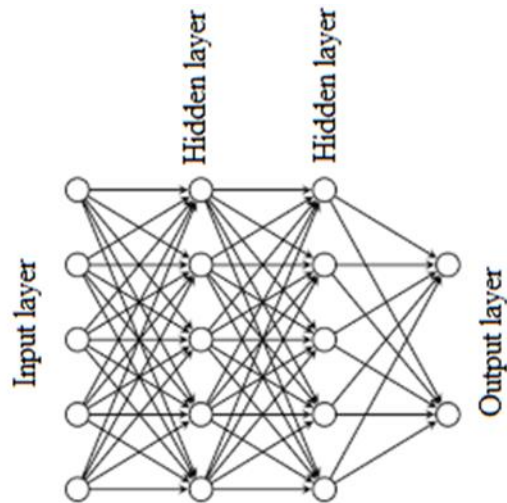


Figure 3.6: Fully Connected layer.

3.2.2 Recurrent Neural Network (RNN)

Strict feedforward structural enterprise fixes not uphold a short-term reminiscence. Any reminiscence belongings are owing to the method historical inputs are re-presented to the network. A humble RNN has activation feedback which exemplifies short-term memory. A national layer is efficient not lone with the outside input of the network but also with activation after the earlier forward propagation. The feedback is adapted by a group of weights as to allow involuntary version finished learning (e.g. backpropagation) [27].

For humble structure and deterministic activation functions, learning can be attained applying alike gradient descent events to those foremost to the back-propagation technique for feed-forward networks. When the activations are stochastic, pretend hardening methods may be additional suitable. The next will appearance at a insufficient of the greatest significant kinds and features of RNN.

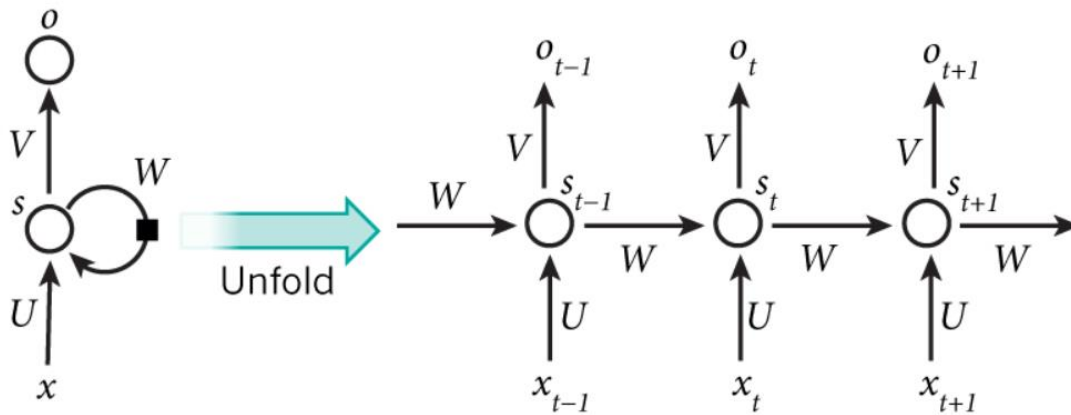


Figure 3.7: RNN Structure.

3.2.3 Long Short Term Memory Networks (LSTM)

An LSTM has an alike switch movement as a RNN. It procedures data transitory on info as it broadcasts forward. The changes are the processes inside the LSTM's cells.

The essential idea of LSTM's are the cell national, and it's numerous entrances. The cell state performance as a conveyance thoroughfare that transfers comparative information all the way depressed the order cable. Gates comprises sigmoid function. A sigmoid function is alike to the tanh function. In its place of squeezing standards between -1 and 1, it squelches standards between 0 and 1.

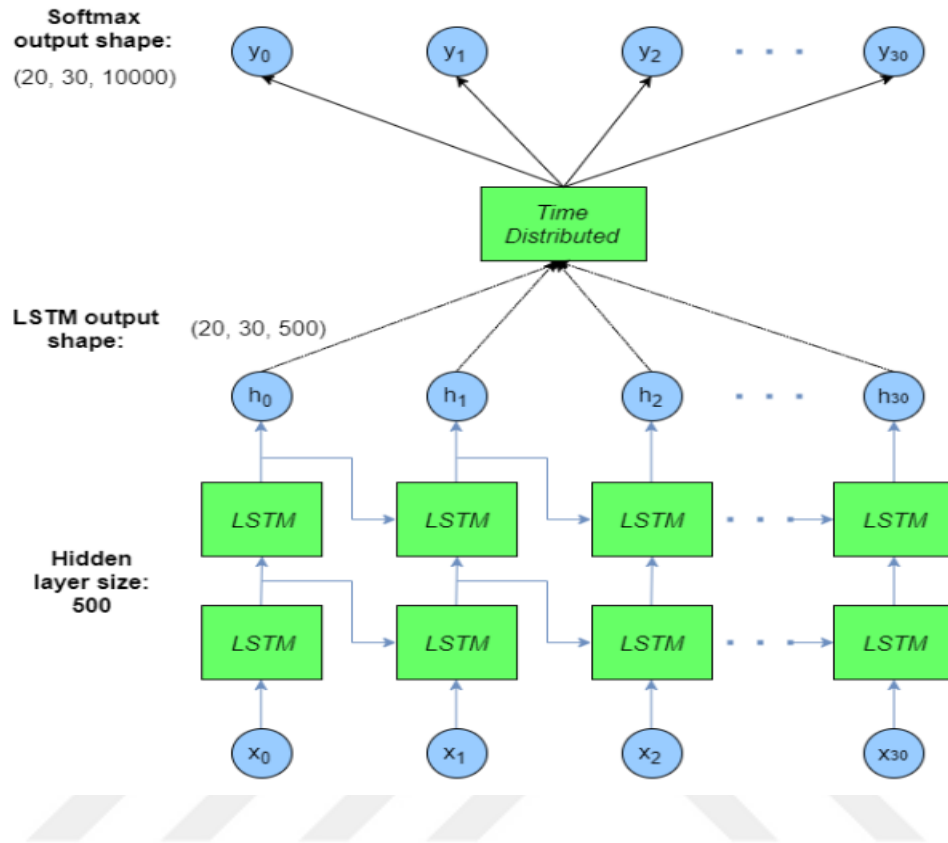


Figure 3.8: LSTM based SoftMax.

Furthermore, structure of LSTM also presented in the Figure 3.9.

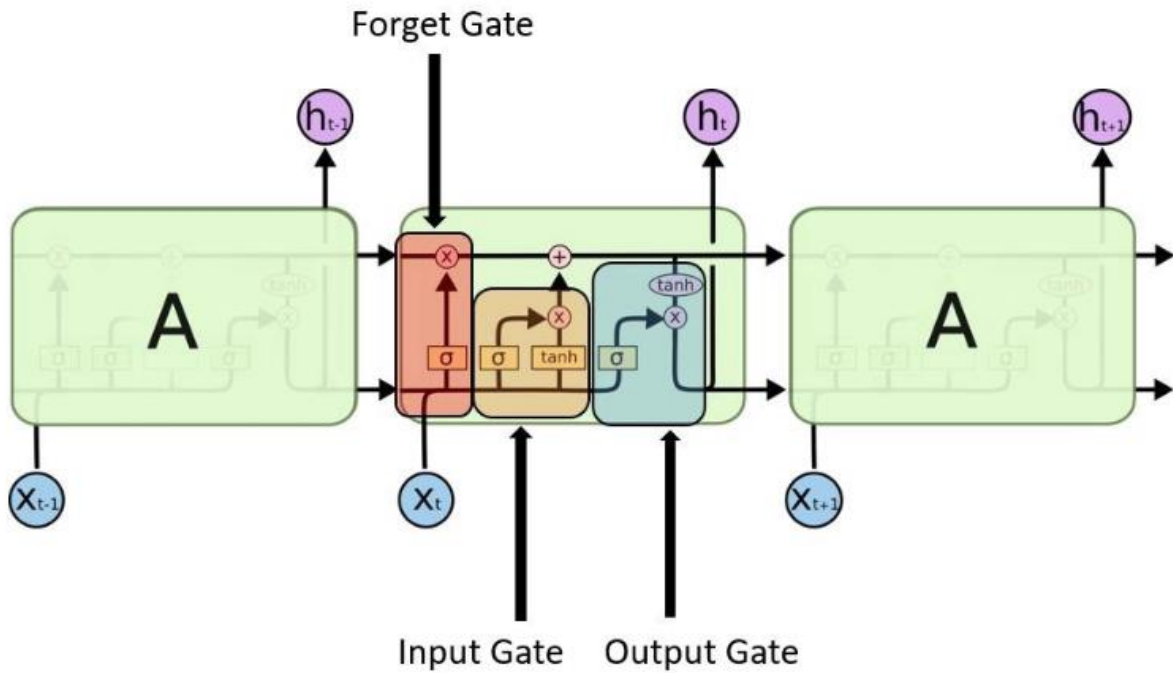


Figure 3.9: LSTM Structure.

3.2.4 Particle Swarm Optimization (PSO) Algorithm

The PSO algorithm is a population-based metaheuristic algorithm proposed by Kennedy and Eberhart, inspired by nature. This algorithm simulates the social behavior of birds to solve problems. The PSO algorithm is created using a series of random solutions that represent particles and you tell your competitors to find the best solution. At each iteration, each particle is modified with two (highest) values, the maximum local and worldwide standards. The first best answer gotten from the particles themselves is displayed with the best local solution and stored as the supposed "best" value. Thus, the other compasses are the best answer derived from all particles in the population; this is the best solution - the global value, more commonly known as the "best" value. Afterward selecting the two best answers, the particles inform their position and velocity applying equations (3.1) and (3.2).

$$X_{k+1}^i = X_k^i + V_{k+1}^i \tag{3.1}$$

$$V_{k+1}^i = wV_k^i + c_1r_1(P_k^i - X_k^i) + c_2r_2(P_k^g - X_k^i) \quad (3.2)$$

Here, X_k^i signifies particle position, V_k^i represents particle velocity, P_k^i represents the best remembered individual particle position (pbest), P_k^g represents the best swarm (gbest), c_1 and c_2 are cognitive and social parameters, r_1, r_2 are chance strictures among (0,1) and w refers inertial coefficient (0,1). This leads to meetings and participation in the PSO process. The PSO method has many advantages over other optimization algorithms. The PSO is easy to justify and has few options to change. In particular, when comparing PSO with the genetic algorithm (GA), which is one of the most important analog algorithms, PSO converges faster and requires fewer parameters for tuning. As a result, PSO has been successfully used in various fields such as neural networks and optimization problems. Algorithm 1 represents the existing PSO algorithm [13].

3.3 PROPOSEDMETHOD

In this section, a new software developed by using CNN for construction cost estimation. The developed software trained by using construction dataset [4] which consist from 169 cases. The dataset divided into training and testing groups 110 for training and 59 for testing. Each case of data analysed by using PSO to reduce the size of the features and the output of PSO plotted as spactogram signal and become input to the CNN as two-dimensional image. The proposed structure of CNN consists from three layers. In the first layer, the convolution layer consists from 8 filters each filter size is 3×3 . Furthermore, max pooling used with two strides. In the second layer, the convolution layer consists from 16 filters each of them consists from 3×3 with max pooling that consist from 2 strides. In the third layer, the convolution layer consists from 32 filters with 3×3 . The max pooling also used in the 3rd layer with two strides. The last layer which is 4th layer that is the output layer that contain the regression function because the target values of our dataset are continuous. The proposed method flowchart presented in figure 3.10.

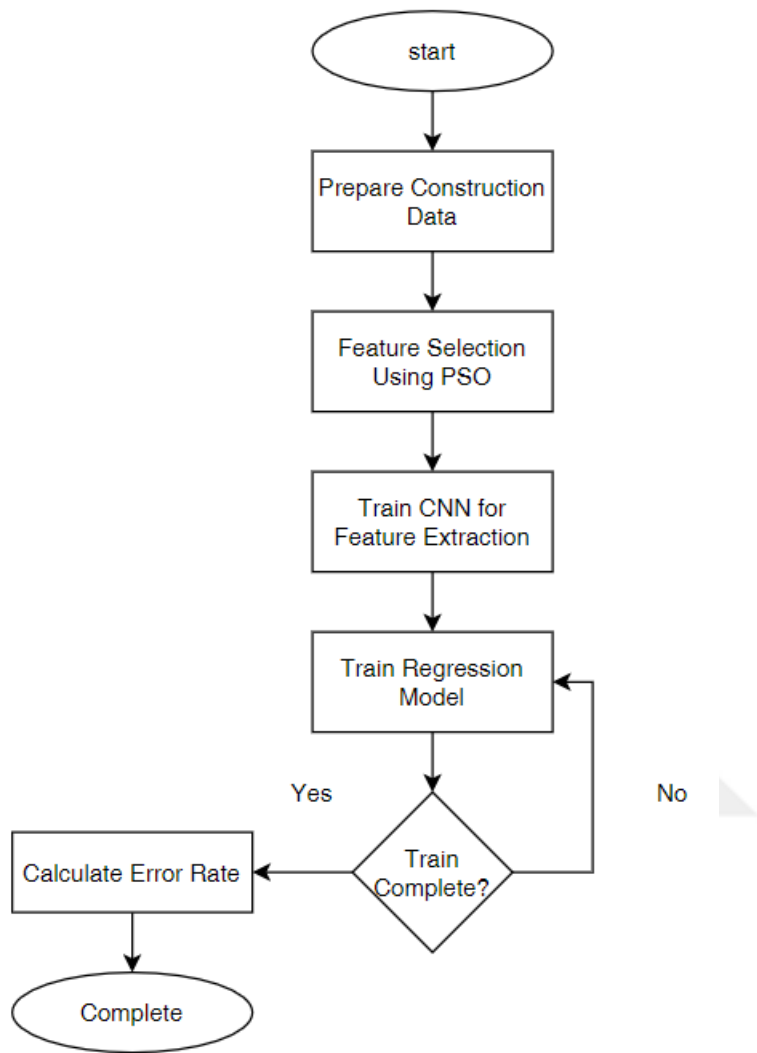


Figure 3.10: Our Cost Estimation Framework.

3.4 EVALUATION PARAMETERS

In this section four parameters are calculated for face recognition problem to evaluate our proposed method the calculated statistical parameters are: Accuracy is how close a measured value is to the actual(true) value.

The confusion matrix is used to evaluate the results of our method by calculating several parameters such as: True positives (TP), True negatives (TN), false positives (FP), and false negatives (FN). The equations (3.3-3.5) represented the evaluation parameters of the model

$$ACC = \frac{(TP + TN)}{(P + N)} \quad (3.3)$$

$$TPR = \frac{TP}{(TP + FN)} \quad (3.4)$$

$$SPC = \frac{TN}{(FP + TN)} \quad (3.5)$$

3.5 VALIDATION USING RANDOM SUBSAMPLING

In this method, random sub-sampling validation is track for a secure amount of K iterations. Throughout apiece iteration it used random sampling minus spare, in instruction to choice a secure amount of S examples that brand up the examination set and are excepted from the training procedure of the perfect. The experiential test statistics are then be around over all iterations. This validation can applied simply and the validation is important to ensure that the model is avoid the overfitting and really its learn from data. The overfitting is very important issue exactly with low size of data and in the CNN technique. There are also several types of validation techniques such as cross folds and else [38, 39].

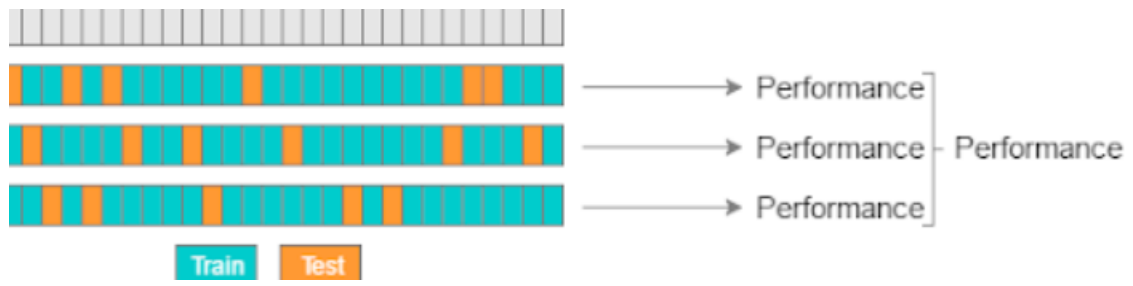


Figure 3.11: Random Subsampling.



4. EXPERIMENTS AND DISCUSSION

4.1 IMPLEMENTATION

The method executed by using MATLAB2020, in the first stage the PSO applied for feature selection. The data obtained and read by using our code the data represented in the figure 4.1.

```
X=[100 2 1 1 2 0 1 0 1 1 1
105 5 3 1 2 0 1 0 1 1 1
111 2 1 1 2 0 1 0 2 1 1
117 2 1 1 2 0 1 0 2 1 1
118 2 1 1 2 0 1 0 2 1 1
125 2 1 1 2 0 1 0 2 1 1
130 5 2 1 2 0 2 0 2 1 1
130 2 1 1 2 0 1 0 2 1 1
135 2 1 1 2 0 1 0 2 1 1
140 2 1 1 2 0 1 0 2 1 1
148 3 2 2 3 0 1 0 2 1 1
150 5 3 1 2 0 3 2 1 1 1
```

Figure 4.1: Read data in MATLAB2020.

The figure 4.1 represented the properties and features of the constructions this data collected from real projects and represented in the form that can be understood by machine learning techniques.

On the other hand, the costs of these features are presented and listed in the figure 4.2.

```

y=[29
77.700
26.740
28.065
29.499
32.075
104.608
33.050
34.710
35.710
94.464
157.100
190.205
170.650
173.900
211.200
185.130
189.150

```

Figure 4.2: Costs of projects.

Then, the The PSO algorithm is a population-based metaheuristic algorithm which used as process of automatically or manually selecting the function that most affects the predicted outcome or outcome of interest. If the data has external characteristics, the model will be less accurate and the model can study external characteristics.

The size of the input features reduced by selecting the important features by using PSO see figure 4.3.

```

nvars = 2;
x = particleswarm(fun,nvars)
fsurf(@(x,y) x.*exp(-(x.^2+y.^2)))

```

Figure 4.3: PSO Script.

Unlimited or constrained derivative-free optimization with PSO solver PSO solves problems with target functions that are not flexible. If pattern search doesn't work as expected, try using it.

After the PSO the features are also read and then plotted and saved, the aim of this step is to convert the data to the images which can be easily understood by CNN. The two dimension data is can easily processed by using with CNN with across of one dimension size of data see figure 4.4.

```

x = [ 34 2 1 1 1 0 1 0 1 1 1
];
x= plot (x);
saveas(x, '1', 'jpg')

x = [ 65 2 2 4 2 0 1 0 2 1 1
];
x= plot (x);
saveas(x, '2', 'jpg')

```

Figure 4.4: Data processing.

Then, the data read from the folders by using `fullfile` which is effective MATLAB2020 used for read list of images and creates a full file specification from the specified file and folder names. Furthermore, Insert platform-specific file separators as needed, but not trailing file separators.

On the other hand, Use the `ImageDatastore` object to manage a collection of image files. Individual images are saved, but not the entire image collection. The `imageDatastore` function lets you create an `ImageDatastore` object, specify its properties, and use the object's functions to get and process data see figure 4.5.

```

fruitstDatasetPath = fullfile('D:\Thesis\Muhanned construction\Training1');
imdsTrain = imageDatastore(fruitstDatasetPath, ...
    'IncludeSubfolders',true,'LabelSource','foldernames');

fruitstDatasetPath = fullfile('D:\Thesis\Muhanned construction\Training1');
imdsValidation = imageDatastore(fruitstDatasetPath, ...
    'IncludeSubfolders',true,'LabelSource','foldernames');

```

Figure 4.5: Read Images using MATLAB2020.

Moreover, the images analysed by using the CNN, which consist from several layers such as input layer, convolutional layer, normalization, pooling and fully connected layers with softmax as transfer learning function. The input layer consist from 656×875 which represented the

dimension of the input images and 3 represented the color which mean these images are not binary images which consist only two bit 0 or 1.

The convolution layer reduces the size of the input image by using sharing weight techniques by extracting sensitive features. In our code, 8 filter with 3*3 size used in the first convolution layer.

Furthermore, maxpooling applied as function which consists from 2*2 sizes with 2 stride sizes. These functions are repeated with the different number of filters but with fixed sizes. The relu layer applied as transfer functions in all layer expect the last layer. The relu convert the pixels that values low than 0 to zero and the positive values remain.

```
layers = [  
    imageInputLayer([656 875 3])  
  
    convolution2dLayer(3,8,'Padding','same')  
    batchNormalizationLayer  
    reluLayer  
  
    maxPooling2dLayer(2,'Stride',2)  
  
    convolution2dLayer(3,16,'Padding','same')  
    batchNormalizationLayer  
    reluLayer  
  
    maxPooling2dLayer(2,'Stride',2)  
  
    convolution2dLayer(3,32,'Padding','same')  
    batchNormalizationLayer  
    reluLayer  
  
    fullyConnectedLayer(1)  
    softmaxLayer  
    regressionLayer];
```

Figure 4.6: CNN layers.

The main purpose of the Softmax function is to express the values of a class as probabilities. Softmax loss is commonly used to solve regression problems. also known as cross-entropy loss. With Softmax function for the first step the unnormalized logarithmic probabilities are calculated by the Softmax function. Then this Non-standardized protocol probabilities are converted to non-

standard probabilities using exponential operation. Then each value is normalized by dividing by the sum value to get the final probabilities.

4.2 RESULTS AND DISCUSSION

In this section our method developed by using MATLAB2018 which random sampling technique applied to the experiments to avoid the overfitting. Five experiments done and finally the average of five experiments presented. In the first stage, each case of project plotted as shown in Figure 4.7. The aim of our software is to use this figure for training and after those other cases for testing. CNN presented remarkable results in computer vision problems and this problem is computer vision problem.

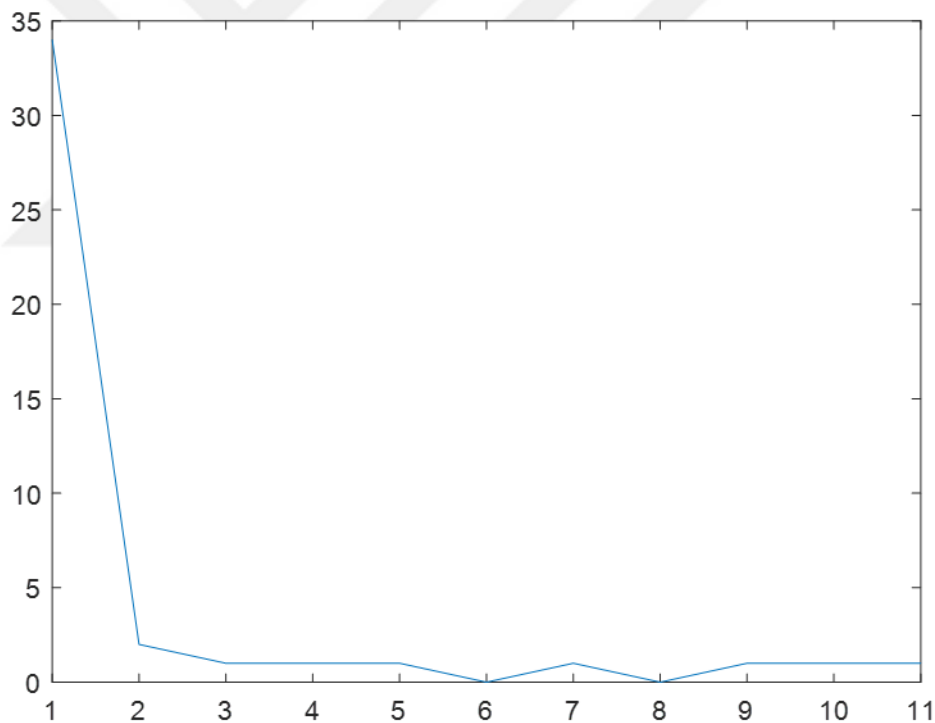


Figure 4.7: Example for Input Data.

The aim of our software is to use this figure for training and after those other cases for testing. CNN presented remarkable results in computer vision problems and this problem is computer vision problem. Then, the 110 cases are used for training our software after training stage complete, 50 cases used for testing our software. The experimental results show that our software presented remarkable results and much closed estimations with actual target see figure 4.8. Furthermore, the summation of actual target is 16707962 \$ and the summation of our software estimation is 16708569 \$. This lead to conclude that our software have ability to estimate very nearest costs to the real target. The error rate in this study is 0, 0018 % which is very low when compared with previous studies.

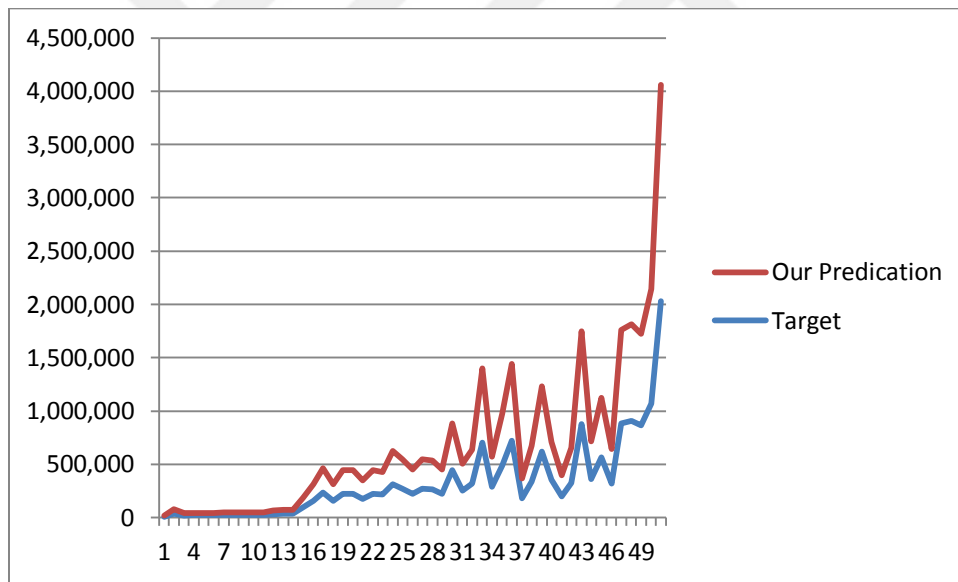


Figure 4.8: Our Software results.

Finally the proposed method compared with two studies proposed in this filed and show that our framework is better when calculated the error rate with these studies, please see Table 4.1.

Table 4.1: Our Software results.

Ref	Model	Error rate (%)
[36]	Neural network	2.3
[37]	Support vector machine	6
Our Method	PSO+CNN+SoftMax	0.0018



5. CONCLUSION

The main conclusions of the study may be presented in this section. In this study, CNN applied to predicate the construction cost estimation, which is new method and first time combined PSO as feature extractor with CNN applied to this problem. The presented technique offered remarkable outcomes when compared with other researches which presented only 0,0018 errors in the total amount when compared with previous studies that have used only SoftMax which presented only % 6 and neural network % 2.3 error rates. As future work, the authors advise to apply various deep learning techniques to this problem to obtain new results. Furthermore, our software can be used with more and new construction data to validate the performance of our method. Finally, other optimization algorithms used as feature extraction and combined with deep learning techniques.

As future work many other techniques can applied to built construction project cost estimation such as deep sparse auto-encoder, LSTM, and RNN. On the other hand, the feature extraction that combined with deep learning techniques also assist the deep learning techniques and improve the estimation rate of the system such as genetic algorithm, gray wolf optimizer and bat algorithm.

REFERENCES

- [1] C. Bortolon, D. Capdevielle and S. Raffard, "Face recognition in schizophrenia disorder: A comprehensive review of behavioral, neuroimaging and neurophysiological studies", 2019.
- [2] S. Farokhi, J. Flusser and U. Ullah Sheikh, "Near infrared face recognition: A literature survey", 2019.
- [3] M. Nappi, S. Ricciardi and M. Tistarelli, "Deceiving faces: When plastic surgery challenges face recognition", 2019.
- [4] Maria Romani, Miriam Vigliante, Noemi Faedda, Serena Rossetti, Lina Pezzuti, Vincenzo Guidetti, Francesco Cardona, Face memory and face recognition in children and adolescents with attention deficit hyperactivity disorder: A systematic review, *Neuroscience & Biobehavioral Reviews*, Volume 89, 2018, Pages 1-12, ISSN 0149-7634, <https://doi.org/10.1016/j.neubiorev.2018.03.026>.
- [5] M. Sugiura, "Three faces of self-face recognition: Potential for a multi-dimensional diagnostic tool", 2019.
- [6] Vinay, A., Gupta, A., Bharadwaj, A., Srinivasan, A., Murthy, K. B., & Natarajan, S. (2018). Deep Learning on Binary Patterns for Face Recognition. *Procedia Computer Science*, 132, 76-83.
- [7] Vinay, A., Cholin, A. S., Bhat, A. D., Murthy, K. B., & Natarajan, S. (2018). An Efficient ORB based Face Recognition framework for Human-Robot Interaction. *Procedia computer science*, 133, 913-923.
- [8] Hamdan, B., & Mokhtar, K. (2018). Face recognition using angular radial transform. *Journal of King Saud University-Computer and Information Sciences*, 30(2), 141-151.
- [9] Vinay, A., BV, S. K., Manoj, P. N., Rao, N. A., Muthy, K. B., & Natarajan, S. (2018). Person Identification in Smart Surveillance Robots using Sparse Interest Points. *Procedia computer science*, 133, 812-822.
- [10] Kautkar, S. N., Koche, R., Keskar, T., Pande, A., Rane, M., & Atkinson, G. A. (2010,

- January). Face recognition based on ridgelet transforms. In *Biometrics Technology* (pp. 35-43).
- [11] Müge Çarıkçı, M., & Özen, F. (2012). A Face Recognition System Based on Eigenfaces Method. *Procedia Technology*, 1, 118-123.
- [12] Mashhoori, A., & Jahromi, M. Z. (2013). Block-wise two-directional 2DPCA with ensemble learning for face recognition. *Neurocomputing*, 108, 111-117.
- [13] Bouzalmat, A., Kharroubi, J., & Zarghili, A. (2013). Face recognition using SVM based on LDA. *International Journal of Computer Science Issues (IJCSI)*, 10(4), 171.
- [14] Yang, J., Zhang, D., Frangi, A. F., & Yang, J. Y. (2004). Two-dimensional PCA: a new approach to appearance-based face representation and recognition. *IEEE transactions on pattern analysis and machine intelligence*, 26(1), 131-137.
- [15] Chen, L. F., Liao, H. Y. M., Ko, M. T., Lin, J. C., & Yu, G. J. (2000). A new LDA-based face recognition system which can solve the small sample size problem. *Pattern recognition*, 33(10), 1713-1726.
- [16] Vinay, A., Vasuki, V., Bhat, S., Jayanth, K. S., Murthy, K. B., & Natarajan, S. (2016). Two dimensionality reduction techniques for surf based face recognition. *Procedia Computer Science*, 85, 241-248..
- [17] Vinay, A., Singh, A., Anand, N., Raj, M., Bharati, A., Murthy, K. B., & Natarajan, S. (2018). Surveillance Robots based on Pose Invariant Face Recognition Using SSIM and Spectral Clustering. *Procedia computer science*, 133, 940-951.
- [18] Congcong Ma, Wenfeng Li, Jingjing Cao, Juan Du, Qimeng Li, Raffaele Gravina, Adaptive sliding window based activity recognition for assisted livings, *Information Fusion*, Volume 53, 2020, Pages 55-65, ISSN 1566-2535, <https://doi.org/10.1016/j.inffus.2019.06.013>..
- [19] Xin Wei, Hui Wang, Bryan Scotney, Huan Wan, Minimum margin loss for deep face recognition, *Pattern Recognition*, Volume 97, 2020, 107012, ISSN 0031-3203, <https://doi.org/10.1016/j.patcog.2019.107012>..

- [20] Mortazavi BJ, Downing NS, Bucholz EM, et al. Analysis of machine learning techniques for heart failure readmissions. *Circ Cardiovasc Qual Outcomes*. 2016;9(6):629–640.
- [21] Keon Myung Lee, Jaesoo Yoo, Sang-Wook Kim, Jee-Hyong Lee, Jiman Hong, Autonomic machine learning platform, *International Journal of Information Management*, 2019.
- [22] Patrick Doupe, James Faghmous, Sanjay Basu, Machine Learning for Health Services Researchers, *Value in Health*, Volume 22, Issue 7, 2019, Pages 808-815, ISSN 1098-3015, <https://doi.org/10.1016/j.jval.2019.02.012>.
- [23] Mohamed El Mohadab, Belaid Bouikhalene, Said Safi, Predicting rank for scientific research papers using supervised learning, *Applied Computing and Informatics*, Volume 15, Issue 2, 2019, Pages 182-190, ISSN 2210-8327, <https://doi.org/10.1016/j.aci.2018.02.002>.
- [24] Yan Cui, Jieli Jiang, Zuojin Hu, Xiaoyan Jiang, Wuxia Yan, Min-ling Zhang, Neighborhood kinship preserving hashing for supervised learning, *Signal Processing: Image Communication*, Volume 76, 2019, Pages 31-40, ISSN 0923-5965, <https://doi.org/10.1016/j.image.2019.04.003>.
- [25] F. Dornaika, Y. El Traboulsi, Joint sparse graph and flexible embedding for graph-based semi-supervised learning, *Neural Networks*, Volume 114, 2019, Pages 91-95, ISSN 0893-6080, <https://doi.org/10.1016/j.neunet.2019.03.002>.
- [26] He Di, Xu Ke, Zhou Peng, Zhou Dongdong, Surface defect classification of steels with a new semi-supervised learning method, *Optics and Lasers in Engineering*, Volume 117, 2019, Pages 40-48, ISSN 0143-8166, <https://doi.org/10.1016/j.optlaseng.2019.01.011>.
- [27] Xiangke Mao, Hui Yang, Shaobin Huang, Ye Liu, Rongsheng Li, Extractive summarization using supervised and unsupervised learning, *Expert Systems with Applications*, Volume 133, 2019, Pages 173-181, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2019.05.011>.
- [28] D. Tian, X. J. Zeng, and J. Keane, “Core-generating approximate minimum entropy discretization for rough set feature selection in pattern classification,” *Int. J. Approx. Reason.*, vol. 52, no. 6, pp. 863–880, 2011.
- [29] B. Liu, Y. Xiao, L. Cao, Z. Hao, and F. Deng, “SVDD-based outlier detection on uncertain

- data,” *Knowl. Inf. Syst.*, vol. 34, no. 3, pp. 597–618, 2013.
- [30] J. Wei, H. Yu, Y. Q. Lu, and J. Wang, “The Research of Bayesian Method from Small Sample of High-dimensional Dataset in Poison Identification,” pp. 705–709.
- [31] R. Kumar, T. Chen, M. Hardt, D. Beymer, K. Brannon, and T. Syeda-Mahmood, “Multiple Kernel Completion and its application to cardiac disease discrimination,” *Proc. - Int. Symp. Biomed. Imaging*, pp. 764–767, 2013.
- [32] C. Li-lin, Z. Hai-chao, Z. Lin-ke, and L. Rui-peng, “Improved k Nearest Neighbors Transductive Confidence Machine for Pattern Recognition ,” vol. 3, no. Iccda, pp. 172–176, 2010.
- [33] D. Tian, X. J. Zeng, and J. Keane, “Core-generating approximate minimum entropy discretization for rough set feature selection in pattern classification,” *Int. J. Approx. Reason.*, vol. 52, no. 6, pp. 863–880, 2011.
- [34] M. Srinivas, R. Bharath, P. Rajalakshmi, and C. K. Mohan, “Multi-level classification: A generic classification method for medical datasets,” *2015 17th Int. Conf. E-Health Networking, Appl. Serv. Heal. 2015*, pp. 262–267, 2016.
- [35] Karim, A. M., Güzel, M. S., Tolun, M. R., Kaya, H., & Çelebi, F. V. (2019). A new framework using deep auto-encoder and energy spectral density for medical waveform data classification and processing. *Biocybernetics and Biomedical Engineering*, 39(1), 148-159. doi:10.1016/j.bbe.2018.11.004.
- [36] Shehatto, O.M. (2013). Cost Estimation for Building Construction Projects in Gaza Strip Using Artificial Neural Network (ANN).
- [37] Rayyan, Tareq M (2017). Cost Estimation of Building Construction Projects in Gaza Strip Using Support Vector Machines Model (SVM). *الجامعة الإسلامية - غزة*. <http://hdl.handle.net/20.500.12358/19086>.
- [38] Guang-Hui Fu, Bing-Yang Zhang, He-Dan Kou, Lun-Zhao Yi, Stable biomarker screening and classification by subsampling-based sparse regularization coupled with support vector machines in metabolomics, *Chemometrics and Intelligent Laboratory Systems*, Volume 160, 2017, Pages 22-31, ISSN 0169-7439,

<https://doi.org/10.1016/j.chemolab.2016.11.006>.

- [39] Qing Wang, Bruce G. Lindsay, Improving cross-validated bandwidth selection using subsampling-extrapolation techniques, Computational Statistics & Data Analysis, Volume 89, 2015, Pages 51-71, ISSN 0167-9473, <https://doi.org/10.1016/j.csda.2015.03.005>.

