Comparative study of Glass Type Classification using Machine Learning

Faizal N[1*], Adarsh G[2]

[1]*Habitat School, Ajman, UAE and University of Kerala, India*
faizalnr@gmail.com
[2]*Department of Computer Science, University of Kerala, India*
adarshg349@gmail.com

Abstract

The glass sector is regarded as one of the world's most important. Glass is used in a multitude of scenarios, ranging from water bottles to X-Ray and Gamma Ray shielding. This is a transparent, non-crystalline, amorphous solid. Glass is used in a variety of ways, and investigators should be able to identify which type of glass is present during a crime scene assessment. We'll utilize the internet dataset and machine learning to resolve the greater than disadvantage in order to find out what type of glass it is. Machine learning algorithms such as with the Support Vector Machine (SVM) algorithm, the Random Forest methodology, the logistical Regression algorithm, and the AdaBoost classifier, as well as the Gradient Boosting classifier, will be exploited. Random Forest performed the best in glass classification out of all the computational rules. Flat glass is now available in a variety of specialized shapes for usage in various applications and products. The process of flotation is used to make flat glass. There are, however, many specific procedures that may be used to give it a range of special forms. The trade is able to meet the needs and desires of the automobile, construction, and solar power industries by using this strategy. However every product and application needs a unique glass, for that reason within the industrial world there square measure differing kinds of glass that may create your domestic and industrial life easier.

Keywords: *Glass Classification, Machine Learning, Principal Component Analysis, K-Nearest Neighbors, Random Forest, Logistic Regression.*

1 Introduction

The glass business is widely acknowledged as one of the most important in the world. From car glass, window glass to water bottles to X-Ray and Gamma Ray shielding, glass is utilized in a wide range of applications. This is an amorphous solid used in variety of industries and that is transparent and non-crystalline. Glass is used in a variety of ways, and investigators should be able to identify which type of glass is present during a crime scene assessment. We'll utilize the internet dataset and machine learning to resolve the greater than disadvantage in order to find out what type of glass it is. Machine learning algorithms such as with the Support Vector Machine (SVM) algorithm, the Random Forest methodology, the logistical Regression algorithm, and the AdaBoost classifier, as well as the Gradient Boosting classifier, will be exploited. Random Forest performed the best in glass classification out of all the computational rules.

Flat glass is now available in a variety of specialized shapes for usage in various applications and products. The process of flotation is used to make flat glass. There are, however, many specific procedures that may be used to give it a range of special forms. The trade is able to meet the needs and desires of the automobile, construction, and solar power industries by using this strategy. However every product and application needs a unique glass, for that reason within the industrial world there square measure differing kinds of glass that may create your domestic and industrial life easier.

Magnesium(Mg), Potassium(K), Barium(Ba), Sodium(Na), Calcium(Ca), Silicon(Si), Iron(Fe), Aluminum(Al) are the elements utilized in the manufacture of several types of glasses (K). In addition, the Refractive Index (RI) is an essential factor in distinguishing between many types of glass and their applications. According to their intended function, the glasses in this dataset are divided into seven categories. Containers, Float glass for building windows, Float glass for vehicle windows, no float glass for building windows, dinnerware, and top lights. A sheet of glass generated by floating molten glass on a liquid metal substrate is termed as float glass. This glass is often of high quality, requires no finishing, and structural flexibility during the production process. This process produces a sheet with a consistent thickness and extremely flat surfaces.

1.1.Objective of the Project

Support Vector Machine (SVM) method, K-nearest neighbors (KNN) formula, Random Forest formula, and Logistic Regression formula are all part of our strategy. Principal Component Analysis (PCA) and other spatiality reduction methods were also employed. We used the Pandas package in Python to examine the dataset and investigate its dimensions. After loading, we tend to use the Matplotlib library to see it. The train and test datasets were then divided by a 5:1 quantitative relationship. After ripping, we usually examine the dataset's choices. Because there are so many possibilities in the dataset, the model will suffer from quality or overfitting. To overcome this limitation, dimensionality reduction methods like as PCA and XGBoost were employed. After a significant dimension reduction, the model was trained and evaluated using a variety of classification methods. We chose the most effective recipe based on its precision. Using this strategy, the industry is able to meet the needs and desires of the automobile, construction, and solar power industries. Furthermore, because each product and use necessitates a different type of glass, the industrial sector offers a variety of options to make your life simpler at home and at work.

1.2.Problems and Issues

This problem has to do with classifying the glass based on its properties. Machine learning approaches such as SVM, Random Forest, or any other classification strategy might be used to tackle this problem. Add extra glass attributes to the model, such as density and heat resistance, to increase the correctness of the proposed system. Now the major issue is dataset is unbalanced and add more features for balancing the dataset.

1.3. Overview of the Work

This disadvantage is attributed to categorizing the glass only on the basis of limited possibilities. This disadvantage is overcome by manipulating machine learning algorithmic rules such as SVM, Random Forest, or other classification algorithmic rules. Because it is a type of proof glass, it is extremely useful as a contact trace material in a broad range of crimes, including robberies and burglaries, hit-and-run accidents, murders, assaults, speed raids, criminal harm, and auto thefts. All of this raises the chance of glass fragments being conveyed to whoever or whatever is accountable when something made of glass breaks. Variations in glass production allow for substantial distinction even with tiny fragments.

To analyze the dataset and investigate its size, then used Python's pandas programme. To determine it during loading, we generally utilize the Matplotlib package. After that, then split the train and used a 5:1 relationship to examine the dataset. It usually assess the dataset's selections after it has been created. Because the dataset contains so many possibilities, the model is likely to suffer from poor quality or overfitting. Spatiality reduction methods such as PCA and XGBoost were employed to overcome this disadvantage. After a significant feature reduction, the model was trained and validated using a variety of classification methods.

1.4. Scope of the Proposed Work

The intended attempt entails utilizing an artificial neural network to introduce a new glass type. According to recent study, 60% of glass cases generated strong proof, with 40% of these cases offering tangible evidence. Depending on the circumstances, the findings may refute the assertion that a person committed a certain crime.

1.5. Organization of the project

The experiment was done out using the JNN Tool and the glass dataset, as previously indicated. In this experiment, a total of 214 samples were utilized. After developing the neural network model with the aid of the JNN coupled to the network layers and analyzing the results on glass classification, as a result, the average inaccuracy in this study's findings has increased. The dataset was then standardized before being uploaded into the JNN environment. The dataset was then divided into coaching and collateral samples, with the number of hidden layers determined.

2 Literature Review

To aid in the criminal investigation, a glass categorization flaw analysis was carried out and whereas if leftover glass is correctly identified, it will be utilized as evidence in the case of a criminal crime. The categorization of glass from the crime scene, and therefore the glass particles recognized as being linked to the crime, is a common need for a procedure. These pieces of glass are generally rather small. It's critical to assess and compare these tiny shards of glass in the context of a rhetorical framework. Each type of glass is made up of a range of components that have completely different (or completely different) unit measurements and refractive indices. The composition and layers of the glass influence the properties of the glass, particularly the refractive index.

Glass categorization hasn't gotten a lot of attention. Mashael S. Aldayel [18] was one of the researchers that explored and tested algorithms like KNN. However, in order to improve the accuracy of a model, we must use additional algorithms. There are several methods for classification; thus, we must choose the best algorithms from the literature study and implement them in Python to categories various eyeglass kinds Vivencio et al [19] proposed utilizing a feature-weighted closest neighbor method with a KNN classifier, based on a chi-square statistical test.

It is time to find out how rare the glass is if the retrieved pieces match the reference sample in terms of RI and chemical composition. This may be done by the scientist using a computer database that contains the combined results of the RI measurements and the chemical analysis findings for each and every reference glass sample that his lab has evaluated. This will tell you how many times a certain type of glass has been observed, but it isn't always a good indicator of its frequency. This is because, despite statistics indicating that glass discovered on clothing is more likely to have come from a container by accident, the bulk of glass samples sent to forensic laboratories come from smashed windows. Aside from that, there are generally tiny pockets of particular types of glass, such as from similar- period buildings. All of this implies that the database's information may be biased and should be handled with caution and safety.

3 Design of glass classification with Machine Learning algorithms

Traditional Identification of the glass at a scene of the crime may be highly helpful in giving evidence in investigations and forensic purposes. Furthermore, because glasses are used in a variety of sectors to make a variety of products, distinguishing the type from its components (oxide content) ahead of time may save money, time, and effort. Because they are a kind of proof glass, they are often used as contact trace evidence in a wide range of crimes, including invasions and robberies, hit and run accidents, murders, assaults, speed raids, criminal damage, and thefts of and from motor vehicles. All of this increases the possibility of glass pieces being transmitted from one type of broken glass to another, whether or not standing was to blame. Even with little fragments, variations in glass manufacturing allow for significant differentiation.

The glass industry has been recognized as one of the world's most important, and glass is used in a variety of applications, from water bottles and window glass to X-Ray and Gamma Ray protection, there's something for everyone. This is an amorphous, non-crystalline material that seems translucent. Glass may be used for a variety of purposes, and investigators should be able to identify which type of glass is present during a crime scene examination. To tackle the greater than disadvantage and we will utilize the net dataset and machine learning to find out the type of glass. Machine learning algorithms such as the Support Vector Machine (SVM) algorithm, the Random Forest algorithm, the logistical Regression algorithm, and the AdaBoost classifier and Gradient Boosting classifier will be used. In terms of algorithmic rules, Random Forest performed the best in glass classification.

3.1. System architecture

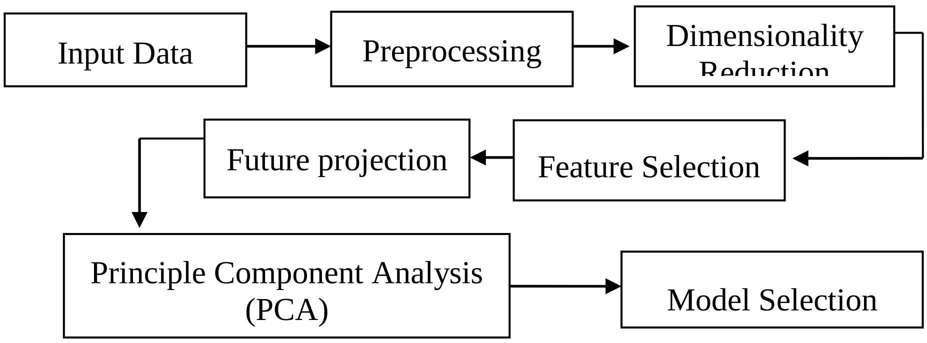

Fig. 3.1 Model System Architecture

3.1.1 Data

The dataset's objective is to forecast Glass based on the provided factors (Id category, RI, Na, Mg, Al, Si, K, Ca, Ba), with all columns save the Id columns playing a vital part in defining the Glass type, which is also our desired variable. A glass dataset with ten input attributes and one output attribute is available (Type of glass). Information on Attributes (Input attributes) Identifier, RI stands for refractive index. Sodium (Na), Magnesium (Mg), Aluminum (Al), Silicon (Si), Potassium (K), Calcium (Ca), Barium (Ba), and Iron (Fe).

3.1.2 Dimensionality reduction

The approach known as spatiality reduction is used to reduce a set of random variables into a set of main variables. It assists the United States of America (USA) in acquiring two dimensional data, allowing us to improve our visual representation of cubic centimeter-based models by producing prediction areas vs prediction border curves for each mode. It eliminates the less significant possibilities and concentrates more on the various alternatives and in our coaching set of knowledge, Dimension Reduction also aids in the removal of superfluous or disordered information occurrences.

3.1.3 Feature selection

Feature selection might be a method of etymologizing a collection of basic alternatives in a variety of ways, depending on the information they provide, accuracy, and prediction mistakes. Dimensionality reduction is the process of reducing a big set of raw data into smaller, more and the manageable groupings for operations. The enormous number of components in these big initial data sets necessitates a considerable amount of processing power. Feature extraction refers to approaches that reduce the amount of data that must be accessed while still correctly and fully characterizing the original data set by choosing and or the combining variables into features.

3.1.4 Feature projection

For reworking the high dimensional knowledge to low dimensional knowledge, it have a tendency to use linear and non-linear reduction techniques based on feature selection. The transformation is based on the concept of a connection between the alternatives in a dataset. During this analysis, the dataset of ten attributes, that area unit all associated with the cell parameters is employed and it have a proclivity to use the PCA method to extract components from a dataset.

3.1.5 Principal Component Analysis (PCA)

The most important component Analysis is an analytical approach for determining the most important alternatives from a datasets of the variance matrix. It reduces a large number of dimensions to just two or three. And it would not be prudent to address the issue of the numerous alternatives right away. PCA is commonly used to minimize the number of variables in knowledge sets with shared variance or other forms of data, such as separate, integrative, ordinal, binary, and symbolic data.

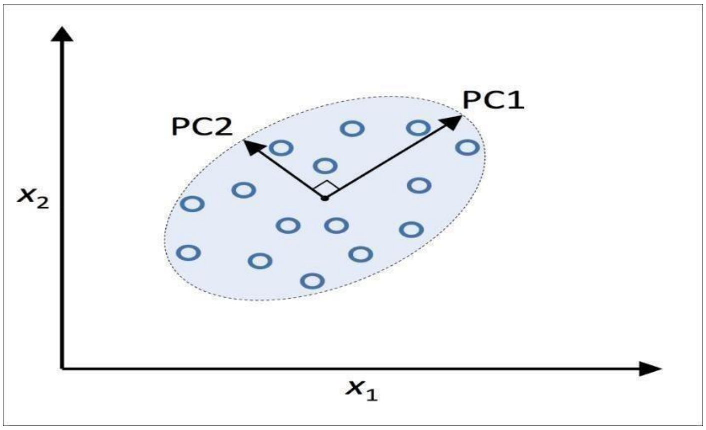

Fig. 3.2 Principal Component Analysis (PCA)

It's a way of condensing information without losing any of the important details from the source. This reduces high dimensions to fewer characteristics in order to maximize variance. For example, x=y produces x autonomous variables are created by combining y autonomous variables. The variance matrix generated for the dataset is used to construct the eigenvectors. The most important portions are the eigenvectors with the highest eigenvalues, which may be used to reconstruct the variance of a huge portion of original data. The dataset is prepared for data processing and machine learning algorithms when PCA is implemented.

3.1.6 Model selection

There are three types of machine learning algorithms such as reinforcement learning, unsupervised learning, and supervised learning. Several pattern recognition difficulties, including as categorization, classification, grouping, and prediction, are overcome using the idea of similarity amid information objects. And it provide specific coaching information to an associated formula that maps input and solves to provide the output in the case of supervised Learning. With the exception of the dataset, the method learns from the data and predicts the outcome. Two supervised learning techniques are regression and classification. Regression is used to predict values in (many or any) situations, Classification, on the other hand, is used to divide data into distinct groups. Unattended learning necessitates the teaching of knowledge, but no input or output mapping is necessary. The formula examines data that hasn't been labelled or categorized and hasn't been given a specific direction. Based on the input, this algorithm produces several types of clusters and then guesses which cluster the data belongs to.

4 Comparison of different Machine Learning algorithms

4.1 Support Vector Machine (SVM)

Each linear SVM as well as kernel SVM are combined in the Support Vector Machine (SVM). SVM rule resolves each regression as well as classification difficulties. Separating knowledge points into distinct categories would be ineffective. In comparison to other algorithms, Support vector machine recognizes complex patterns in advanced datasets with ease. The most important goal of SVM is to maximize the distance between knowledge points and the hyperplane

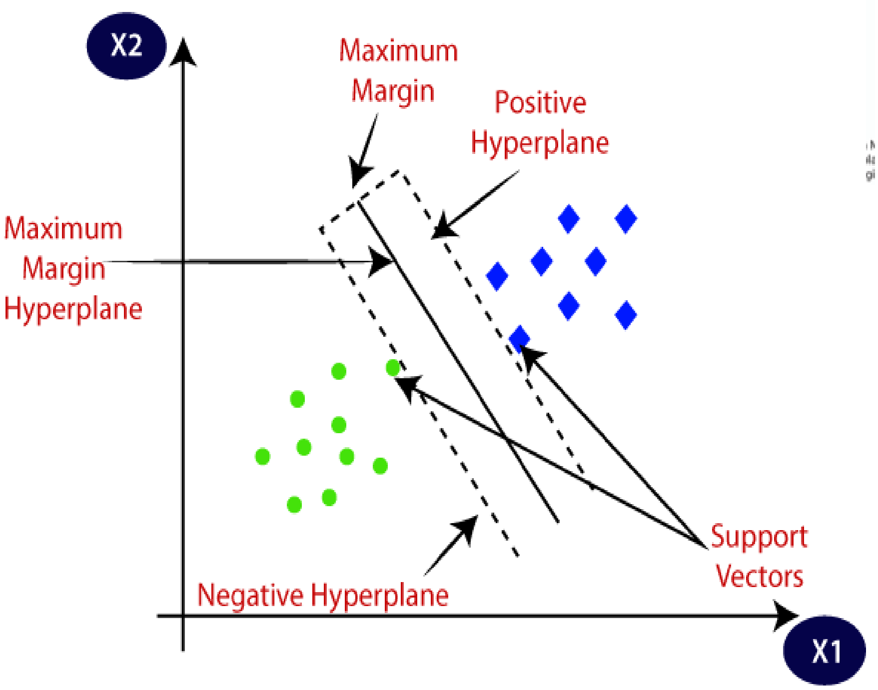

Fig. 4.1 Example of SVM classifier

SVM is most commonly used to solve classification issues. Allow us to suppose that n is the number of alternatives we have in our knowledge in order to create a plot. And It has a proclivity to express each goal as a knowledge portion in a very n-dimensional space with a coordinate function price for each characteristic. They had a proclivity to classify by using the hyper plane that divides the two categories from one another, which supported the alternatives. With the use of hyperplanes, the chosen boundaries are categorized. Observant knowledge points on diverse aspects of the plane can reveal entirely new categories of points. The count of characteristics determines the size of hyperplanes.

4.2 K-Nearest Neighbors (KNN) Algorithm

The K Nearest Neighbor formula, often known as the KNN formula, is a simple formula that employs the whole dataset as part of its coaching portion. When a prediction is required for an unknown knowledge instance and the system searches the whole coaching knowledge set for the k-most comparable instances, and then returns the data with the most similar instance as the forecast. The number K in the k nearest neighbor formula denotes the number of nearest neighbor points that voted for the new check data category. The KNN formula will address both classification and regression problems.

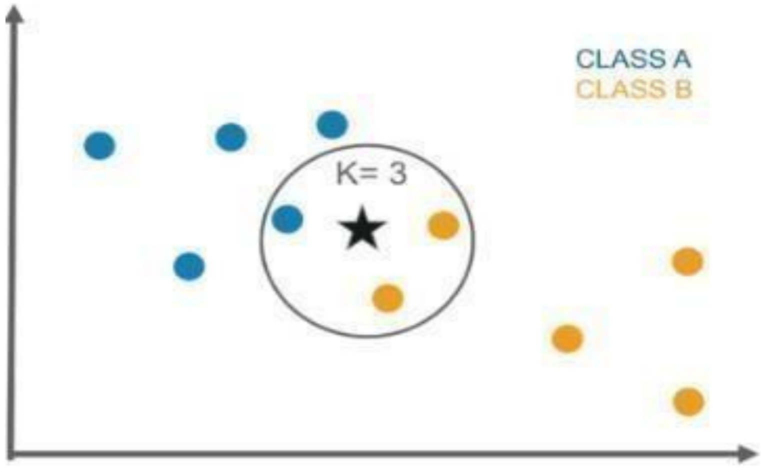

Fig. 4.2 Example of KNN classifier

KNN works by calculating the distances between a query and each data sample, selecting the required number of instances (K), and then voting for the most frequent label or averaging the labels (in the case of regression). They utilized KNN throughout the study and obtained the least reliable findings.

4.3 Random Forest Algorithm

The Ensemble machine learning formula, also known as Random Forest, is a method of merging multiple classifiers to tackle a difficult issue. It predicts the ultimate result of the data using a majority selection approach. The prediction categories have been chosen from among the trees. The number of trees included in the model determines its accuracy.

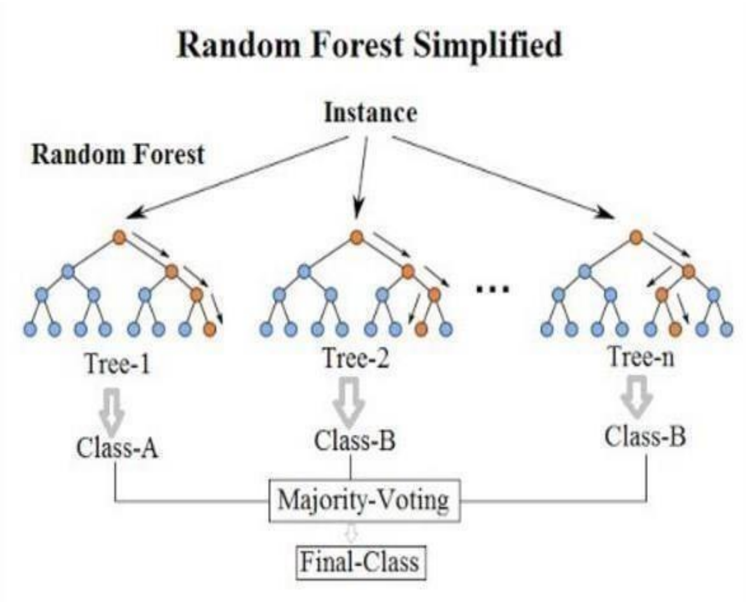

Fig. 4.3 Random forest classifier

It also takes care of the problem of overfitting. The decision tree is the basic building block of random forest classifiers. A decision tree is a hierarchical structure created from a data set's characteristics (or independent variables). The decision tree is divided into nodes based on a measure connected with a subset of the characteristics. The random forest is a set of decision trees linked to a set of bootstrap samples derived from the original data set.

4.4 Logistic Regression

Providing Regression is a supervised milliliter algorithmic rule similar to linear regression, but instead of regression, For classification, logistic regression is utilized, and multivariate analysis is used to look at the relationship between one split variable and one or more independent variables (categorical or continuous). It depicts the freelancing variables' linear connection and categorizes them into binary categories.s

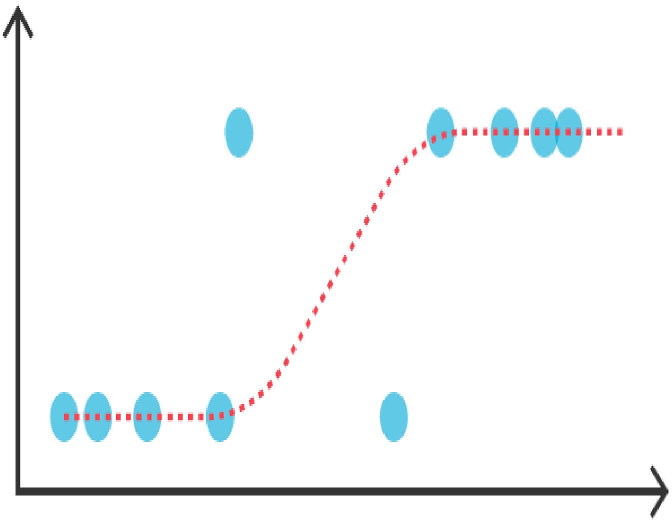

Fig. 4.4 Logistic Regression

4.5 Neural[10] Network

A neural network is a collection of algorithms that aims to recognize underlying links in a large amount of data in a way that replicates how the human brain works. The biological structure generated by the operation of the human nervous system is known as a neural network.

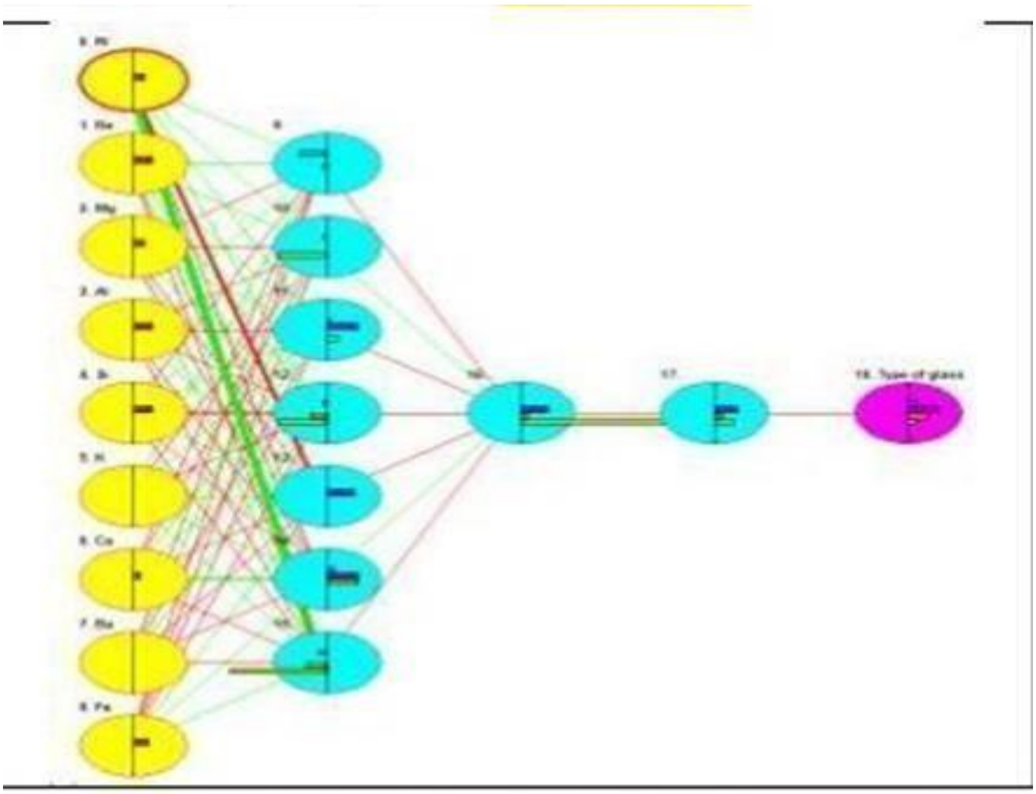

Fig. 4.5 Neural Network model

Because of its brain capacity, which allows it to extract rules and learn from data neural networks are being utilized in a wide range of applications to construct a network model that may be used for data classification, pattern recognition, and prediction.

4.6 Artificial Neural Network

The human brain is impressed by Artificial Neural Network algorithms. The factitious neurons are linked and interact with each other. Each association is weighed by prior learning experiences, and each new piece of knowledge adds to the learning process. It's a deep learning algorithmic software that's inspired by the way a biological brain works. And these are multilayer structure with an associated input layer, an associated output layer, and several hidden levels.

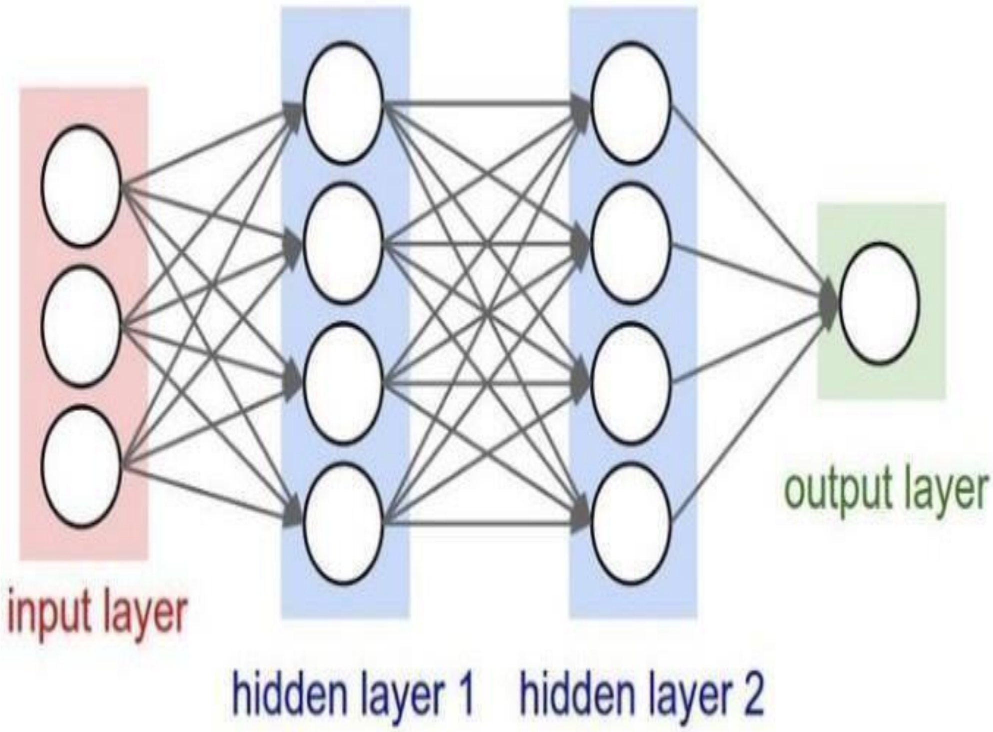

Fig. 4.6 Artificial Neural Network

A layer is made up of numerous neurons that are connected to one another. A non-linear transformation operator (sigmoid function) links the signal received from the previous layer's neurons to a response signal supplied to the next layer's neurons in each nerve cell. There is forward transmission as well as backward transmission in this Artificial Neural Network is a term used to describe a type of neural network that is made from Because of its capacity to generalize and resistance to clangorous and erroneous information, artificial neural networks are becoming increasingly popular these days. A great deal of research is being done to improve the efficacy and accuracy of neural specification modelling and coaching.

Despite the fact that a great deal of research has been done, there is currently no comprehensive theory addressing the precise range of information size, network architecture, and the optimum algorithmic program for ANN modelling, as the choice of those parameters is dependent on the information's character.

5 Result Analysis

Let will construct and examine the glass classification using machine learning methods in this chapter, much as we did in Chapter three. The results of studies of the quantity and types of glass shards discovered on rich people's clothing, accessories, covers, and consumer items show that discovering a high number of glass shards from a steady supply on someone's clothing entirely out of the blue is unusual.

5.1 Dataset

Id number	RI	Na	Mg	Al	Si	K	Ca	Ba	Fe	Type of glass
1	1.52101	13.64	4.49	1.1	71.78	0.06	8.75	0	0	1
2	1.51761	13.89	3.6	1.36	72.73	0.48	7.83	0	0	1
3	1.51618	13.53	3.55	1.54	72.99	0.39	7.78	0	0	1
4	1.51766	13.21	3.69	1.29	72.61	0.57	8.22	0	0	1
5	1.51742	13.27	3.62	1.24	73.08	0.55	8.07	0	0	1
6	1.51596	12.79	3.61	1.62	72.97	0.64	8.07	0	0.26	1
7	1.51743	13.3	3.6	1.14	73.09	0.58	8.17	0	0	1
8	1.51756	13.15	3.61	1.05	73.24	0.57	8.24	0	0	1
9	1.51918	14.04	3.58	1.37	72.08	0.56	8.3	0	0	1
10	1.51755	13	3.6	1.36	72.99	0.57	8.4	0	0.11	1

Fig 5.1 Dataset sample

Type of glass (output attributes) Float-processed windows are used in construction. Building windows that have not been float treated, float processed vehicle windows and Containers, dinnerware, and headlamps are all common household goods.

	RI	Na	Mg	Al	Si	K	Ca	Ba	Fe	Type
0	1.52101	13.64	4.49	1.10	71.78	0.06	8.75	0.0	0.00	1
1	1.51761	13.89	3.60	1.36	72.73	0.48	7.83	0.0	0.00	1
2	1.51618	13.53	3.55	1.54	72.99	0.39	7.78	0.0	0.00	1
3	1.51766	13.21	3.69	1.29	72.61	0.57	8.22	0.0	0.00	1
4	1.51742	13.27	3.62	1.24	73.08	0.55	8.07	0.0	0.00	1
...	...	...	...	...	...	...	...	...	...	...
145	1.51839	12.85	3.67	1.24	72.57	0.62	8.68	0.0	0.35	2
146	1.51769	13.65	3.66	1.11	72.77	0.11	8.60	0.0	0.00	3
147	1.51610	13.33	3.53	1.34	72.67	0.56	8.33	0.0	0.00	3
148	1.51670	13.24	3.57	1.38	72.70	0.56	8.44	0.0	0.10	3
149	1.51643	12.16	3.52	1.35	72.89	0.57	8.53	0.0	0.00	3

Sample data of the actual dataset

Fig 5.2 sample data

It includes 214 distinct varieties of glasses that may be classified depending on their intended use.

5.2 Experimentation

Neural networks are extensively utilized in a range of applications, owing to its mentality. They can extract rules from data and learn from it to create a network model that can be used for classification, pattern recognition, and data prediction. The most promising aspect of the Neural network that other categorization approaches lacked is that it aids in the simulation of the network and the creation of a model that can be used more frequently and applied to fresh data that hasn't been exposed to the network previously. Just-NN is a tool that may be used to model neural networks. The JNN tool aids in the creation of a neural network model to coach and evaluate data for categorization, discovery of hidden patterns, grouping, and future prediction.

The experiment was carried out using the JNN Tool with the glass dataset as described above, and the dataset had 214 samples during the experiment. The dataset is usually split into 154 training samples and 60 validation samples in this study. It used the JNN to construct a neural network model that they looked at the results of the categorization of the kind of glass as well as the average error discovered within the findings, which were connected to the layers of the network.

It explains how input parameters are related to hidden layers and how further hidden layers are connected to the output layer in the model that was developed.

The total weights of the connections from the input node to any or all of the nodes inside the first hidden layer make up the input significance of a neural network. Designers usually normalize the dataset before uploading it to the JNN environment. Researchers usually choose the number of hidden layers, divide the dataset into workable chunks, and confirm samples.

5.3 Experimental Results

The total weights of the connections from the input node to any or all of the nodes inside the first hidden layer make up the input significance of a neural network. Usually normalize the dataset before uploading it to JNN. We typically choose the number of hidden layers, divide the dataset into workable chunks, and confirm samples.

Principal Component Cumulative Variance

De-correlate the characteristics that are linearly dependent using a PCA, and then plot the cumulative explained variance. The total weights of a neural network determine its input significance of the connections from the input node to all or any or any of the nodes at intervals of the first hidden layer. Typically usually normalize the dataset before uploading it to the JNN environment. Then chose the number of hidden layers and split the dataset into employment and confirmatory samples

14

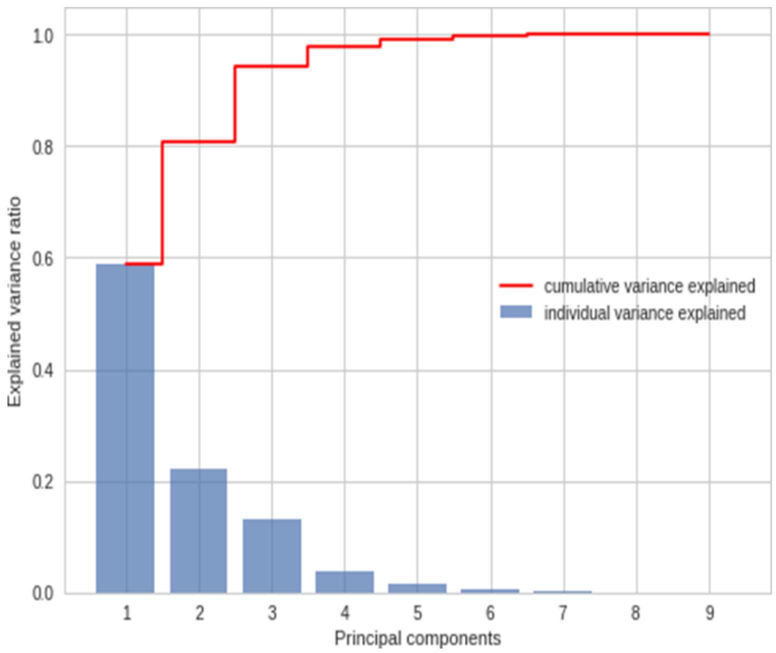

Fig 5.3 Cumulative variance of PCA

Diagnose the performance of the algorithms

Plot the learning and validation curves to diagnose overfitting, then fine-tune and finish the model.

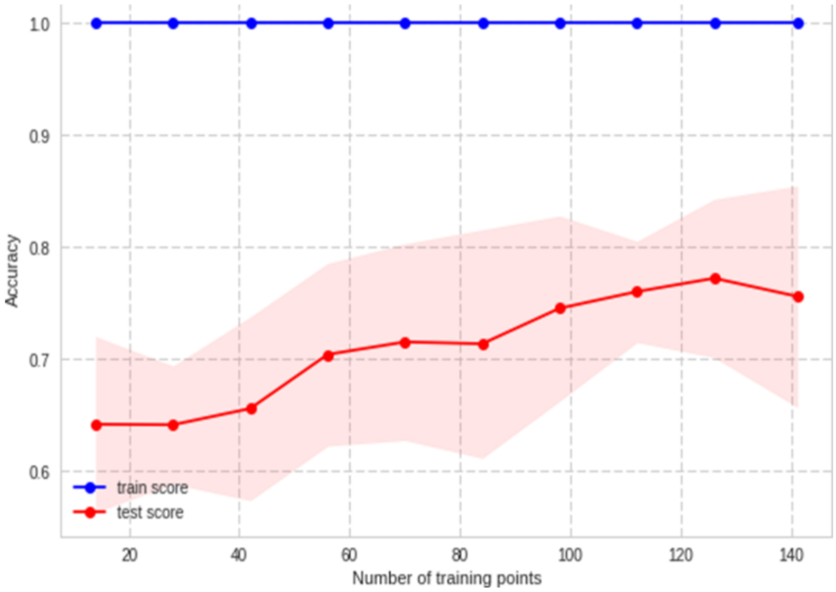

Fig 4.4 Learning curve of the model

5.4 Results Discussion

This experiment employed a dataset of seven different types of glasses, each of which had nine characteristics and one output. Based on element measurements and refractive indices, utilized a variety of machine learning algorithms to determine the kind of glass. It used five distinct machine learning algorithms. The Support Vector Machine (SVM) method, the K-Nearest Neighbors (KNN) technique, the Random Forest algorithm, the Artificial Neural Network (ANN), and the Logistic Regression algorithm are all examples of machine learning algorithms. With 79.62 percent accuracy, the model based on the Random Forest method fared best, followed by Support Vector Machine (SVM) with 77.77 percent accuracy.

6 Conclusion and Future Works

The authors utilized a variety of machine learning techniques to forecast the glass class. Based on the quality of the algorithm model, the best model is determined which is suited to address this sort of issue. For the classification purpose a dataset with nine characteristics and one class output that included seven distinct types of spectacles. Based on element measurements and refractive indices, the model is applied by a variety of machine learning algorithms to determine the kind of glass.

As can be seen from the data, the average error reduces the number of times we train our network model grows as the number of times to train it grows. Overtraining, on the other hand, may result in a higher rate of mistake. ANN is a classification system for different kinds of glass. In this study used five distinct machine learning algorithms. The SVM method, the Artificial Neural Network (ANN), the K-Nearest Neighbors (KNN), the Random forest algorithm, and the Logistic Regression algorithm are all examples of machine learning algorithms.

Future work will be done to refine the model and make it more useful in real-world situations. Improve the suggested system's accuracy by include more glass qualities in the model, such as density and heat resistance. The main difficulty now is that the dataset is imbalanced, and new features are needed to balance the dataset.

Acknowledgments. The authors are thankful to the Department of Computer Science, University of Kerala, India, for the facilitation and support given for the completion of the given study.

References

1. B.German 2019). UCI Machine Learning Repository []vCentral Research Establishment Home Office Forensic Science Service Aldermeston. https://archive.ics.uci.edu/ml/datasets/Glass+Identification
2. Osisanwo F.Y., Akinsola J.E.T., Awodele O., Hinmikaiye J. O., Olakanmi O., Akinjobi J. "Supervised Machine Learning Algorithms: Classification and Comparison". International Journal of Computer Trends and Technology (IJCTT) V48(3):128-138, June 2017. ISSN:2231-2803. Published by Seventh Sense Research Group. DOI: http://dx.doi.org/10.14445/22312803/IJCTT-V48P126
3. Ibrahim, Nagaraju Kolla & M. Giridhar Kumar. Supervised Learning Algorithms of Machine Learning: Prediction of Brand Loyalty. International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278- 3075, Volume-8 Issue-11, September 2019. DOI: 10.35940/ijitee.J9498.098111
4. S Aruna and Dr S P Rajagopalan. Article:A Novel SVM based CSSFFS Feature Selection

Algorithm for Detecting Breast Cancer. *International Journal of Computer Applications* 31(8):14-20, October 2011.

5. Chao-Ying Joanne Peng, Kuk Lida Lee & Gary M. Ingersoll (2002) An Introduction to Logistic Regression Analysis and Reporting, The Journal of Educational Research, 96:1, 3-14, DOI: https://doi.org/10.1080/00220670209598786
6. Smith HM. Interpreting Qualitative Data: Methods for Analyzing Talk, Text and Interaction 3rd Edition. Sociological Research Online. 2006;11(4):100-101. DOI: https://doi.org/10.1177/136078040601100403
7. Devroye, L., Gyorfi, L., Krzyzak, A. & Lugosi, G. (1994) "On the strong universal consistency of nearest neighbor regression function estimates", Ann. Statist, 22: 1371– 1385. https://www.jstor.org/stable/2242230
8. Devroye, L. & Wagner, T.J. (1977) "The strong uniform consistency of nearest neighbor density estimates", Ann. Statist., 5: 536–540. https://www.jstor.org/stable/2958905
9. Mohammed, Somia B., Ahmed Khalid, and SaifeEldin F. Osman. “A Survey of Classification Methods.” International Journal of Advanced Engineering Research and Science 3.10 (2016): 148–152. Web.
10. Jolliffe IT. 2002. Principal component analysis, 2nd edn New York, NY: Springer-Verlag. https://doi.org/10.1007/b98835
11. Bernhard Flury. 1988. Common principal components & related multivariate models. John Wiley & Sons, Inc., USA..
12. Marc Hallin, Davy Paindaveine & Thomas Verdebout (2014) Efficient R-Estimation of Principal and Common Principal Components, Journal of the American Statistical Association, 109:507, 1071-1083, https://doi.org/10.1080/01621459.2014.880057
13. Schapire, R. and Freund, Y. Boosting: Foundations and Algorithms. MITPress, 2012.
14. Liu Zhongying, Zhao Qinghua, (Bidding for construction projects)2007,224- 225
15. Mashael S. Aldayel. 2012. K-Nearest Neighbor classification for glass identification problem. ICCSII.2012.645452
16. D. P. Vivencio, E. Hruschka, M. Nicoletti, E. dos Santos, and S. Galvao, “Feature-weighted k-Nearest Neighbor Classifier,” Foundations of Computational Intelligence, 2007. FOCI 2007. IEEE Symposium, pp. 481– 486, 2007.