[1]*Suryakant V.Patil
[2] Renuka R. Londhe

A Review on Multi Factor Dependent Prediction and Analysis through Machine Model

Journal of Electrical Systems

Abstract: - In areas like automation, speech recognition, and vision, the proliferation of machine learning (ML) techniques offers excellent potential for predicting and identifying the system. It is difficult to predict and analyse outcomes that depend on several variables, and doing so has important ramifications for many different fields. The current state of research on multi-factor dependent prediction and analysis using machine models is thoroughly reviewed in this review. It examines the strategies and techniques used to deal with the challenges posed by modelling the interdependencies between various components. The article compares machine models to conventional statistical approaches in analysing multi-factor interdependence, highlighting both their benefits and drawbacks. It looks at how ML algorithms can be used to solve difficult prediction problems with multiple variables. The review also highlights how crucial feature engineering and selection are to raising prediction accuracy. It examines methods for extracting useful characteristics from various data sources so that the models may accurately reflect underlying trends and connections.

Keywords: multi-factor; prediction; detection; machine learning; emotion detection

I. INTRODUCTION

Humans have utilised a range of tools since the dawn of time to accomplish a variety of tasks more swiftly. The inventiveness of the human brain led to the development of several inventions. Humans benefit from these robots' ability to meet a number of needs, including those related to calculating, travel, specific industries, and ML, which is among them. Arthur Samuel referred to ML as the field of study that enables computers to learn without being formally instructed. For its checker playing programme, Arthur Samuel gained notoriety. ML has been employed to train machines to manage information more effectively. Even after examining the data, people may struggle to evaluate and extrapolate insights from it. In that situation, apply ML with humans. The abundance of databases has boosted the demand for machine learning. Additionally, numerous companies employ ML to collect relevant data. The objective of ML was to gain knowledge from data. Several experiments have attempted to teach robots how to study on themselves without being properly programmed. It requires handling vast volumes of data, and many programmers and mathematicians utilise a number of ways to handle this problem [1]. Information retrieval, speech identification, human-computer communication, data mining, visualisation, and computer graphics are just a few of the many disciplines where machine learning (ML) was effectively used. Nevertheless, due of its illogical functionalities and murky operating principle, majority users frequently treat an ML framework as a "black box" [2]–[4]. High-performance algorithms are often developed through an effort-intensive trial-and-error approach lacking a clear knowledge of the reasons why the framework operates. In order to better comprehend and analyse ML frameworks, particularly their internal working processes, academic investigators and industrial operators are confronted with issues that call for more accessible and explicable systems. More specifically, ML has been described as a collection of techniques that can autonomously find trends in data and utilise those trends to forecast future data or make other types of uncertain decisions. Unpredictability in ML can take many different forms, such as: what constitutes the most accurate future forecast given some previous data, or what was the most appropriate algorithm to clarify certain information, what assessment should I take next, etc. Although closely connected to data, the probabilistic method to ML varies slightly from it with regard to vocabulary and focus. Clustering or Unsupervised ML, Semi-Supervised ML or Reinforcement Learning like Dynamic Programming and Markov Chains, and Supervised ML like Regression and classification constitute some of the most significant ML techniques [5].

Additionally, ML aids in the discovery of fixes to numerous issues in automation, speech identification, and vision. Let's use the ability to recognise faces as an instance: This seems to be a simple job to do; regardless variances in position, illumination, hairstyle, and other factors, people regularly recognise relatives and friends by

[1] *Corresponding author: College of Computer Science and IT, Latur, SRTMU, Nanded, Maharashtra

[2] Author 2 Affiliation: College of Computer Science and IT, Latur, SRTMU, Nanded, Maharashtra

simply observing their facial features or through images. However, humans have been difficult to clarify how people do it because individuals do it instinctively. People have been unable to express their skills, hence the researchers are unable to develop the computer programme. Humans are aware that a person's face has shape and that a facial image was not merely an arbitrary set of pixels. It appears to be balanced. The face parts that are located in precise places include the nose, mouth, and eyes. Every person's face has a distinctive design formed up of a particular configuration of features. A training programme may recognise an individual by looking for a characteristic unique to that individual in a particular image after analysing example face photographs of that individual. One illustration of pattern identification was this [6]. Furthermore, computers are programmed using instances of data or prior knowledge in ML to optimise the effectiveness criterion. A framework that was constructed to a certain extent is already in the hands of the researchers, and learning entails using a software programme that optimises the model's parameters using training data or past information [5]. The model may be explanatory so that knowledge can be gained from the data or predictive so that future predictions may be made. The idea of statistical was applied in the development of logical models because the main goal of ML had been to extract implications from instances. Computer science serves two purposes. In order to tackle the optimisation issue and preserve and handle the enormous quantity of data people typically have, investigation first need effective algorithms for training. Secondly, a model's structure and mathematical solution for interpretation must be effective once it has been learnt. The effectiveness of a training or inferred method, specifically its dimension and temporal complexity, might be just as crucial in some situations as the forecast accuracy [6]. The type of method used relies on the issue type, the number of parameters, the type of framework that would work best, and other factors. The present research reviews the multi factor dependent like pattern identification, speech recognition, visualization, etc. using different ML algorithms.

Rest of the section: Section 2 describes the recent literatures related to forecasting and analysis employing ML models; Section 3 reviewed the literatures on machine models employed for multi factor dependent detection and prediction; Section 4 provides a research question and its results; Section 5 reveals the analysis of performance and the discussion of assessment and Section 6 explains the conclusion.

II. REVIEW OF LITERATURE

Some of the recent articles related to prediction and analysis using ML models are described as follows:

Speech emotion identification is now a significant field of study in the fields of signal processing. Emotion voice recognition has been backed by a number of helpful technologies. Muljono et al. [7] looked into how well the mel-frequency based cepstral coefficients (MFCC) recognised emotions in Indonesian speech. Moreover, the database consists of audio files from Indonesian films that were downloaded from a website. To separate the movie's sound, certain preprocessing is done. Anger, sadness, happiness, and neutrality have been the 4 emotion categories into which the audio recordings have been divided. The Indonesian language's emotional content can be recognised using support vector machines (SVM). The most popular classification and feature extraction techniques in voice recognition have been SVM and MFCC. On a number of SVM kernel-based functions, including the linear kernel, radial basis function (RBF), and polynomial kernel, the MFCC performance has been contrasted. According to the findings, SVM employing a linear kernel yields 66% as maximum accuracy, while SVM employing a polynomial kernel generates only 45%. However, the computational time of the model is high.

In order to solve the issue of forecasting cricket match outcomes using previous game data from the Indian Premier League (IPL), Kumash et al. [8] looked into ML technologies. Filter-based approaches such as Information Gain (IG), Correlation-oriented attribute selection, ReliefF, and Wrapper were employed to find the dataset's important characteristics. A more significant development is the adoption of ML methods, such as Random Forest (RF), K-Nearest Neighbour (KNN), Model Trees (classification using regression), and Naïve Bayes (NB), to produce prediction algorithms from separate feature sets obtained by the filter-predicated approaches. The toss decision and the home team benefit have been used to create two emphasised subsets. To create a predictive approach, specific ML approaches have been used to both attribute sets. Experimental studies revealed that when contrasted with statistical and probabilistic approaches, tree-based frameworks, notably RF, functioned better. Nevertheless, nothing of the ML techniques under consideration produced good predictive frameworks on the Toss emphasised subset.

Li-ion battery prediction has been a difficult problem that restricts the development of technology like electric cars and consumer electronics. Electrochemical impedance spectroscopy (EIS), non-intrusive, a real-word, and statistics-rich test that has previously been underutilised in battery assessment, has thus been combined with

Gaussian process ML to create a precise battery prediction mechanism by Yunwei et al. [9]. The biggest archive of this type that the researchers are aware of includes commercial Li-ion batteries, which is more than 20,000 EIS spectra recorded at various temperatures, stages of impose, and health conditions. Without additional analysis of features, the given Gaussian process approach autonomously detects which spectral properties forecast degradation from the source spectrum. Furthermore, despite not having a thorough understanding of the battery's previous operational circumstances, the model makes an accurate prediction of its remaining usable life. The outcomes showed how important EIS signals are for battery management platforms. However, the model has higher computational time.

A crucial aspect of contemporary company intelligence was sales prediction. The hard work of sales prediction is made more difficult by a lack of information, inadequate numbers, or abnormalities. Regression has proven more of a problem for sales prediction than time sequence. Scientists may utilise supervised ML algorithms to find complex patterns among the sales kinematics that also contain a lot of risk factors. For a business to succeed, its sales estimates must be accurate. Businesses can recognise potential risks and make better decisions by using a solid revenue projection system. In order to analyse the Rossmann sales data, Shreya Kohli et al. [10] used predictive techniques including KNN and linear regression. An accurate sales prediction can help a business cut costs related to surplus stock, make informed decisions about the future, and increase profitability. Therefore, it is essential to evaluate the model using statistical methods such MAPE and RMSE. To establish which classification algorithm has performed better at forecasting sales, the results are used. The model isn't tested on a variety of databases, though.

Because it has a high death rate worldwide, heart disease has become a critical health problem for many people. A crucial objective is to identify cardiovascular issues using routine medical data analysis, especially coronary artery diseases, heart attacks, etc. Early identification of cardiac issues can save many lives. The application of ML can aid in both accurate forecasting and decision-making. The use of ML techniques in the medical field is developing considerably. Kavitha et al. [11] therefore suggested a unique ML technique to forecast cardiac disease. The Cleveland heart illness database was utilized in the presented research, and data mining methods including regression and classification were applied. RF and DT ML methods have been used. The ML model's innovative model has been created. Three ML algorithms—RF, DT, and an integrated approach (a combination of DT and RF)—are employed in the execution. According to experimental findings, the hybrid algorithm's heart disease forecasting accuracy was 88.7%. The connection is made to solicit user data in order to forecast cardiac illness, for which the study employed a DT and RF combined model. However, the forecasting accuracy needs more enhancement.

III. Review on Multi Factor Dependent Prediction and Detection Using Machine Model

A ML framework for automated emotion recognition from speech was developed by Kholodna et al. [12]. The created model will be employed to the framework for tracking public sentiment. The article has offered a brief examination of existing research works on the procedure for developing ML systems for automated emotion recognition from speech. Consideration has been given to both traditional and deep ML techniques, as well as certain initial database characteristics. A built ML algorithm for automated emotion identification from text was the outcome of the study. The deep learning (DL) approach, whose architecture combines multiple levels of bidirectional long-short term memory (Bi-LSTM) using a pre-trained word embedding fastText framework, has been selected for experimentation with real-series data gathered from the statistics resources. The Daily Dialogue and its potential for instructing the classificatory were taken into consideration. It has also been discussed how to create and choose the best model for automatically detecting emotions in speech. The ML algorithm produced acceptable results after being created and enhanced through experimentation with real-world information. The constructed model had issues correctly classifying "noisy" data, hence the analytical findings acquired with it were unable to be regarded accurate. A method called EmotexStream was created by Maryam et al. [13] to automatically detect emotions in the flow of text publishes on social media sites such as Twitter. 90% of recordings are accurately classified by the model, according to experiments. In order to determine the likelihood that a record corresponds to a certain emotional group, the researchers also address the issue of fuzzy boundaries across emotional classes. They do this by employing a dimensional framework for emotions and fuzzy categorization. Additionally, Rahul et al. [14] created an experimental framework for social media short comments and postings emotion identification. The study made use of the CNN - BiLSTM, LR, CNN-LSTM and NB classifier with

fastText vectorization standard techniques. The CNN-LSTM approach came in second, while the LR approach performed the best.

According to research by Harsh et al. [15], guided waves (GW) produced by surface-bonded piezoelectric wafer transducers (PWT) can be used to detect high-frequency damage to rails. In order to comprehend the guided wave's propagation properties and how they deal with head defect in a tiny rail sample, an integrated practical and simulation investigation is first given. Furthermore, an approach built on a ML algorithm has been suggested to effectively find damage to rail heads. Utilising the signal's frequency, time-frequency-domain, and time properties, an ML model is learned. The suggested framework has been then put to the test utilising simulation and experimental data from randomised rail head impairment. The range of the severity estimation error was determined to be between 2.00% and 16.67%. An effective approach of rail damage identification termed SCueU-Net was presented by Jun Lu et al. [16]. For the primary time, the challenge of rail damage identification uses an integration of the U-Net graph segmentation structure and the saliency cues approach of damage localization. The experimental findings demonstrated that the method considerably increases the rail damage identification effectiveness, with 99.76% detection accuracy rate, which was 6.74% greater than the recently developed approach in damage recognition accuracy. Using a temporal sequence of UAV images, Saba et al. [17] established a reliable and automatic method for determining the soybeans relative maturity. The sequential behaviour of time-series information is captured by an end-to-end integrated framework that combines CNN and LSTM. Tests of the suggested DL model were conducted across the US in six different contexts. Results indicated that the recommended CNN-LSTM framework was superior to the local regression approach in terms of efficiency. The researcher also shows how this newly discovered knowledge might be applied to help with judgements on plant breeding development.

In order to incorporate football domain information into the modelling process, Daniel et al. [18] suggested two innovative ideas. Two novel feature engineering techniques for match result forecasting were created by the investigator according to these concepts. They are called rating feature learning and recency feature extraction. Researchers created two learning groups from the Challenging data employing these techniques. The KNN framework learned on the score feature learning batch finished first in the 2017 Soccer Forecasting Challenge. With a combination of extreme gradient boosted trees (XGBoost), researchers could marginally enhance on this outcome in subsequent studies. According to the research, the effective integration of domain information into the ML modelling process is a crucial component in the forecasting of football match results. A groundbreaking comparison research for horticultural sales projections using nine cutting-edge ML as well as three traditional approaches was presented by Florian et al. [19]. The study has been predicated on conventional retail data from the horticultural industry, which had particular features, such as size and periodicity. To replicate the efficient functioning of a prediction system with an ongoing data updating and a possibly shifting information distribution, the researcher used a prediction setup with a frequent refit of the algorithm's variables. Moreover, the outcomes demonstrated the supremacy of ML techniques, particularly XGB. This advantage grew with bigger databases with several seasons. Furthermore, researchers reported improved performance when they added extra features like weather or vacation information. Lastly, the study demonstrated the computational efficiency of the top performance XGB.

Plant diseases reduce agricultural output, which affects the economy. As a result, forecasting techniques for identifying plant diseases and evaluation must be created. If caught early enough, the most prevalent disease, fungus infection, can be treated by adopting the proper precautions. Manish et al. [20] designed an expert mechanism for forecasting some of the fungal illnesses like powdery mildew, root rot/leaf blight, anthracnose, and rust. For the diseases categorization, a multi-layered perceptron (MLP) approach is utilised, which not solely successfully identifies plant diseases yet may also dramatically enhance productivity. The suggested method includes the dataset pretreatment, exploratory assessment of data, and identification component as three key components. Plant disease prediction has been done using the suggested ML strategy. The experimental findings showed that the model works better with regards to accuracy than a number of current techniques. Over 98% of predictions for each disease turned out to be accurate on average. This research establishes the viability of employing this method for less expensively and more quickly identifying plant diseases. Farmers will benefit from the early identification of plant diseases to prevent further losses. Kshyanaprava et al. [21] concentrated on supervised ML algorithms including RF, DT, KNN, NB, and SVM for maize crop disease identification using the plant images. To choose the most accurate model for predicting plant diseases, the aforementioned classification

strategies are examined and contrasted. Comparing the RF method to the remaining categorization methods, it achieves the best accuracy of 79.23%. Producers will employ all of the aforementioned trained models for the prompt identification and categorization of emerging image illnesses as a preventative step. A system for ML-based rice leaf illness identification was presented by Kawcher et al. [22]. This research identifies bacterial leaf blight, leaf smut, and brown spot illnesses as 3 of the foremost prevalent diseases affecting rice plants. The input consisted of crisp images of damaged rice leaves on a white backdrop. After the database underwent the appropriate pre-processing, a variety of ML techniques, comprising KNN, J48(DT), NB, and LR, were trained on it. With different levels of accuracy, the models forecasted the illnesses affecting rice leaves. On testing data, it has been discovered that DT functioned most effectively, with accuracy of 97.91%.

ML algorithms may be employed to anticipate stock movements because financial news and social media offer information that can alter investor behaviour. Wasiat et al. [23] apply techniques on financial news and social media data to determine how this data affects the stock market forecasts accuracy for 10 days in the future. Spam tweet minimization and feature selection have been carried out on the databases to enhance predictions quality and performance. Researchers also conduct tests to identify stock markets, which have been challenging to forecast and those that have been most affected by financial news and social media. Researchers compared the output of various methods to identify a reliable classifier. Lastly, deep learning (DL) is employed for maximising forecasting accuracy, and certain classifiers have been ensembled. The experimental findings demonstrated that the RF-based classifier remains unified, and its ensemble achieved 83.22% accuracy. Waqas et al. [24] adopt a unique DL-based technique for anticipating electrical load. The implementation of a three-phase model also included the use of an integrated feature selector (XGboost and DT), a feature extraction method (Recursive Feature Elimination), and upgraded Extreme Learning Machine (ELM) and SVM for categorization and prediction. The Grid Search (GS) method was used to modify the SVM's hyperparameters, while the Genetic Algorithm (GA) has been used to optimise the ELM's hyperparameters. The simulation outcomes have been displayed in graphs, and the data are supplied in tabular form. It is evident from these findings that the offered enhanced techniques perform more accurately and effectively than cutting-edge methods. SVM-GS and ELM-GA had prediction accuracies of 93.25% and 96.3%, correspondingly. The reported SVM-GS and ELM-GA algorithms are more accurate than contemporary methods by 7% and 10%, correspondingly. Sayma and Muhammad [25] have suggested a novel ML categorization method for PCOS forecasting. The algorithm was learned and examined on ultrasonography images of 594 ovaries, and for extracting attributes from the gathered images, CNN incorporating various cutting-edge methods and transfer learning had been utilised. On that lowered dataset, stacking ensemble ML method employing traditional algorithms as initial learners and boosting or bagging ensemble method as meta-learner were then applied. Contrasting the suggested method with other ML-based methods now in use, the accuracy has been greatly improved whereas training time for implementation has also decreased. The greatest outcomes, once again using the suggested extended method, have been achieved by combining the "VGGNet16" pre-trained method with CNN structure as an attribute extractor and stacked ensemble system with the "XGBoost" framework as a meta-learner for classifying images; the model has attained 99.89% accuracy for categorization.

Table.1: Overall summary of different model

References	Method	Domain	Merits	Demerits
[12]	ML-BiLSTM	Emotion identification	The procedure for developing and choosing the optimum model for automatic emotion recognition in speech is covered in the paper. This includes things like choosing features, fine-tuning hyperparameters, and performance assessment measures. For both academics and professionals attempting to create comparable platforms, such insights have been helpful.	The developed model struggled to categorise "noisy" data accurately, making it impossible to view the analytical results it produced as accurate.

[13]	Supervised ML	Emotion identification	The suggested method enables the algorithm to discover patterns and connections between textual characteristics and emotions, producing precise categorization outcomes.	The specifics of the database utilised for both training and assessment are not stated in the paper. The generalizability and dependability of the produced models can be strongly impacted by the dataset's dimensions, variety, and representativeness.
[14]	LR	Emotion recognition	The LR model is fine-tuned using custom variables, enabling greater optimisation and adaption to the particular issue area.	The testing set operates badly and is unable to capture generalised features. The main problem seems the learning set's data being unbalanced.
[15]	SVM	Damage detection	This method has the ability to detect problems early, enabling prompt maintenance and averting further degradation or incidents.	The size and variety of the rail specimens utilised in the trials are not specifically mentioned in the study. The adaptability of the suggested technique to actual rail systems might be impacted by a small or restricted sample size.
[16]	SCueU-Net	Damage identification	This fusion of methods enables a thorough and efficient analysis of rail photos, increasing the precision of damage detection.	The paper makes no specific mention of the suggested method's computational effectiveness.
[17]	CNN-LSTM	Maturity prediction	This method enables precise estimation of soybeans comparative maturity by effective analysis and interpretation of the time ordering of UAV images.	The research doesn't specifically address the computing needs and sustainability of the suggested framework.
[18]	KNN, XGBoost	Outcome prediction	These methods provide creative ways to get pertinent data from the given data and include it in the modelling procedure.	Interpreting DL and ensemble models, like XGBoost, may be difficult and perplexing.
[19]	9 ML models	Sales prediction	The study illustrates the practical difficulties and difficulties associated with sales predictions	The comprehensibility of the ML

			in the horticultural industry employing real-world statistics. This improves the research results' credibility and usefulness.	algorithms used in the investigation is not covered in the investigation.
[20]	MLP	Plant disease forecast	The suggested strategy can help increase agricultural output by precisely diagnosing plant illnesses. The transmission of illnesses can be stopped and agricultural losses can be reduced with prompt identification and appropriate safeguards. Increased agricultural yield could result from this, which would lessen the economic harm caused by illnesses.	Predicting specific fungal illnesses remains the main focus of the research. It might not offer a complete remedy for other kinds of plant illnesses brought on by various infections. To apply the strategy to a larger variety of disorders, additional study and refinement might be required.
[21]	RF	Maize disease recognition	Producers may use the trained algorithms created through the study to quickly identify and classify newly emergent image-based plant illnesses. Farmers may identify infections early thanks to this preventive method, halting their spread and minimising crop damages. Appropriate management can increase agricultural output and lower financial losses.	The research only uses image analysis to identify diseases, which might have some drawbacks. Simple eye inspection simply might not be sufficient to recognise some diseases' mild signs. The accuracy and dependability of disease identification could be improved by including additional data resources, like environmental variables or genetic indicators.
[22]	DT	Leaf disease identification	Crisp images of damaged rice leaves taken against a white background serve as the system's input. With this method, data collecting is made simpler and integration into useful agricultural technologies is made simple. The research offers a non-invasive and immediately perceptible method for identifying illnesses by concentrating on visual aspects.	The consistency and calibre of the source images have a significant impact on how well the disease detection system performs. The ML models performance might be impacted by poor image quality or by a wide range of image situations. For

				actual implementation, robustness to various image circumstances should be taken into account.
[23]	RF	Stock market forecasting	The study deals with the problem of spam tweets and applies strategies to lessen their influence on forecasts. The study improves the quality and dependability of the data employed for assessment by removing irrelevant or deceptive information. The most pertinent features have been also chosen using feature selection approaches, which can enhance the prediction models performance.	Scalability and real-time execution of the suggested approach are not discussed in the paper. Real-time processing of massive amounts of financial news and social media data can be computationally challenging. Future research should take into account the strategies' viability and scalability in real-world situations.
[24]	SVM-GS, ELM-GA	Electricity load prediction	The suggested strategy improves the effectiveness and efficacy of load forecasting algorithms by choosing the most pertinent features and lowering the complexity of the database.	Gathering data, real-time forecasting, and interaction with current systems are examples of practical execution considerations that are not included in the research. When using load forecasting algorithms in realistic situations, these factors are essential. By taking these factors into account, the methods suggested would be more practical and useful.
[25]	XGBoost	PCOS forecasting	The study makes use of an ML stacking technique that combines conventional ML methods as initial learners and boosts or bags ensemble methods as a meta-learner. By utilising each learner's skills and minimising bias, stacking ensembles offer the possibility to enhance model effectiveness. This method makes it possible to classify PCOS more	Among the study's shortcomings is the difficulty of gathering sufferer clinical images in the lowest developed nations, which led to the application of ML algorithms to a small sample size

			precisely using ultrasound images.	of images because of an absence of a database.
[26]	SVM with linear kernel	Autism forecasting	SVM-based classifier has performed higher and attained higher efficiency	Additional non-verbal factors are required to increase categorization accuracy and find possible cross-domain correlations because the sample size employed for classification remained small.
[27]	9 ML methods	Autism forecasting	Nine ML algorithms have been created, and their effectiveness in distinguishing between people with ASD and those without it was assessed; the NN classifier produced the greatest reliability in the binary categorization.	In this research, PCD data heterogeneity has not been extensively explored. Gender discrimination in data collecting is acknowledged by the authors.
[28]	Multilayer NN named CGRNN	Autism prediction	Higher accuracy was attained for both classification and feature extraction	The sample size is too tiny to make any trustworthy conclusions
[29]	LR, RF, and KNN	Autism forecasting	LR method has attained higher performance with sensitivity 100% and specificity 96.59	There has been no discussion of the logic behind the efficiency of LR. There has been no description of the important characteristics or their function in ASD.
[30]	SVM	Asperger Syndrome (AS) identification	By employing SVM classifiers, the research harnesses the strengths of this algorithm to achieve accurate AS prediction.	The study had some restrictions, including the smaller sample size used in the analysis, which might limits the finding's generalizability.
[31]	Gradient Boosting Classifier (GBC), RF, Recursive Feature Elimination (RFE)	AS recognition	The study leverages the power of ML to automate the disease identification process. ML algorithms can handle complex patterns and variations in data, enhancing the accuracy of AS predictions.	Larger datasets are needed to validate the findings and improve the generalizability of the method.
[32]	SVM, RFC, NB,	Attention-Deficit/Hyp	By exploring various models, the research provides insights into the	The biggest impediment to the

	LR, and KNN	eractivity Disorder (ADHD)	performance of different approaches and identifies the most effective classifier.	success of this research seems to be the absence of large, open-source ASD databases.
[33]	DeepFMRI	ADHD identificatio n	By selecting the most relevant features and reducing the dimensionality of the dataset, the proposed method enhances the efficiency and performance of prediction models.	The minimal data size getting analyzed is among the suggested work's shortcomings.

IV. RESEARCH QUESTIONS

The review's objectives are to identify the various state-of-the-art ML-based detection and prediction approaches as well as future research directions to match the capabilities of current methods. Therefore, two Research Questions (RQs) were developed to achieve those goals:

1. Which ML methods are being commonly used for multi dependent factor prediction?
2. What are the several ML methods considered for multi dependent factor detection?
3. What is the effectiveness of using ensemble methods in detection/prediction process?

Research question results:

The most popular ML method for detecting multi-factor dependence makes use of XGBoost [18], [19] and LR [14] methods. Since then, the ML model has garnered a lot of attention and respect. It also has the ability to handle a considerable volume of data. The study's focus can shift from prediction and identification to classification as a result of how effectively machine learning models pick up the necessary detailed properties from the image data itself. The ML techniques beat conventional methods, demonstrating their capacity to recognise intricate patterns and boost prediction precision. This demonstrates the potential of ML to improve forecasting abilities across several industries.

The most often used ensemble technique for identifying multi-factor dependence uses XGBoost-based techniques [18], [19] and [25]. Since it can continuously obtain deep features from images by adapting to small variations in the images, XGBoost have gained a lot of praise and reputation in the field of machine learning (ML). Additionally, it has the capacity to manage a considerable amount of data. Additionally, researchers currently prefer using ML techniques over traditional strategies due to their success in detection and prediction, notably ensemble methods. Because an ensemble learning method efficiently learns the pertinent intricate characteristics from the image data itself, the study's emphasis could change from feature extraction and reduction to network structural layout. The results showed increased end-to-end learning dependability.

V. PERFORMANCE ANALYSIS AND DISCUSSION

The accuracy of every model is taken into account while performing performance studies for various recognition and prediction approaches. Accuracy represents the proportion of correctly identified image pixels. This is also referred to as full pixel precision. Despite representing the most fundamental performance indicator, if there has been a category conflict, the outcomes of image recognition may be incorrect. When one of the specified categories performs better than the other, there is a class discrepancy. In this scenario, findings would be skewed as a result of the superior accuracy of the dominant class surpassing the inferior reliability of the opposing class. If there was no group distinction, the accuracy metric has been proposed for evaluating detection outcomes employing images.

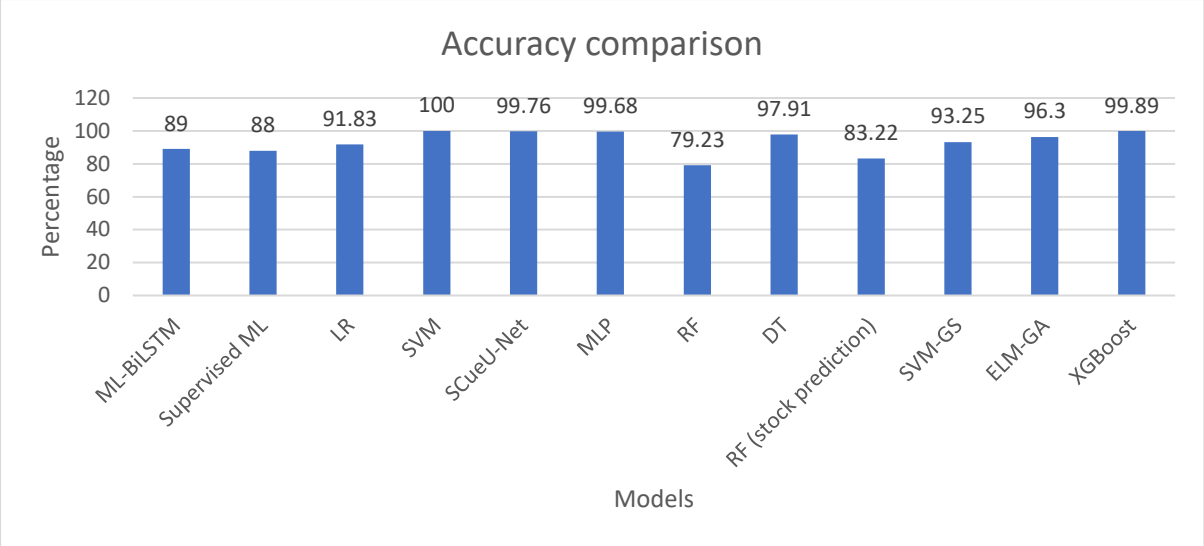

Figure 1:Accuracy comparison

The models selected for comparison are ML-BiLSTM [12], Supervised ML [13], LR [14], SVM [15], SCueU-Net [16], MLP [20], RF [21], DT [22], RF-based stock market prediction [23], SVM-GS and ELM-GA [24], and XGBoost [25]. The result indicated that SVM [15] model has attained higher accuracy while compared to other DL and ML-based methods. In Figure 1, the performance of several models is displayed. SVM works effectively even in high-dimensional spaces, which makes it appropriate for issues involving numerous characteristics. It enables precise predictions and detections by capturing complicated linkages and decision limits in high-dimensional data. SVM has built-in resistance to overfitting, which happens when a model is overly complicated and matches the training samples too closely, resulting in subpar generalisation to new data. In order to obtain the most generalizable decision boundary and lower the likelihood of overfitting, SVM aims to locate the ideal hyperplane that maximises the margin between classes. Support vectors, or samples close to the decision limits, have been the sole part of the training information that are used in SVM. Due to its ability to concentrate on the most useful cases while ignoring the less informative ones, SVM is able to handle big datasets with efficiency. Compared to methods that take into account the complete dataset, this results in quicker training and forecasting times.

The prediction research would be extended to include image classification in addition to noisy image reduction. Neural network methods will boost the training database and produce more usable features, improving classification accuracy. More intensity categories along with geometrical conditions will be included in the techniques for model-based forecasting. Making the best decision for a particular application may be difficult. As a consequence, it is possible to combine a variety of approaches to achieve the essential goal of categorization, forecasting, and recognition. Therefore, utilising hybrid methodologies could aid future researchers in overcoming the drawbacks of each individual method. In a hybrid method, a classifier that tends towards non-linear categorization will have its performance enhanced by altering a few kernel functions.

VI. Conclusions

The review of multi-factor dependent prediction and analysis using machine model underlines the value and promise of using ML models to complicated prediction issues involving numerous components. In order to shed light on the numerous methodologies and strategies used to solve multi-factor dependencies such emotion detection, damage identification, maturity, sales, and outcome prediction, the review analyses the body of current literature and research works in this field. The article highlights the benefits and drawbacks of using ML models to identify and analyse the relationships between various parameters, which traditional statistical methods could find difficult to do. A variety of data sources and factors can be used into ML models like XGBoost and SVM to increase prediction accuracy. Researchers can now concentrate on creating new hybrid algorithms that are specifically made to tackle multi-factor dependent prediction problems in the future. These algorithms ought to be able to capture intricate connections and interactions between several parameters with greater accuracy and performance.

REFERENCES

[1] B. Mahesh, "Machine learning algorithms-a review," *Int. J. Sci. Res. IJSRInternet*, vol. 9, pp. 381–386, 2020.

[2] M. Liu, J. Shi, Z. Li, C. Li, J. Zhu, and S. Liu, "Towards Better Analysis of Deep Convolutional Neural Networks," IEEE Trans. Vis. Comput. Graph., vol. 23, no. 1, pp. 91–100, Jan. 2017, doi: 10.1109/TVCG.2016.2598831.

[3] T. Muhlbacher, H. Piringer, S. Gratzl, M. Sedlmair, and M. Streit, "Opening the Black Box: Strategies for Increased User Involvement in Existing Algorithm Implementations," IEEE Trans. Vis. Comput. Graph., vol. 20, no. 12, pp. 1643–1652, Dec. 2014, doi: 10.1109/TVCG.2014.2346578.

[4] J.-D. Fekete, "Visual Analytics Infrastructures: From Data Management to Exploration," Computer, vol. 46, no. 7, pp. 22–29, Jul. 2013, doi: 10.1109/MC.2013.120.

[5] A. C. Müller and S. Guido, Introduction to Machine Learning with Python: A Guide for Data Scientists. O'Reilly Media, Inc., 2016.

[6] E. Alpaydin, Introduction to Machine Learning, fourth edition. MIT Press, 2020.

[7] Muljono, M. R. Prasetya, A. Harjoko, and C. Supriyanto, "Speech Emotion Recognition of Indonesian Movie Audio Tracks based on MFCC and SVM," in 2019 International Conference on contemporary Computing and Informatics (IC3I), Singapore, Singapore: IEEE, Dec. 2019, pp. 22–25. doi: 10.1109/IC3I46837.2019.9055509.

[8] K. Kapadia, H. Abdel-Jaber, F. Thabtah, and W. Hadi, "Sport analytics for cricket game results using machine learning: An experimental study," Appl. Comput. Inform., vol. 18, no. 3/4, pp. 256–266, Jun. 2022, doi: 10.1016/j.aci.2019.11.006.

[9] Y. Zhang, Q. Tang, Y. Zhang, J. Wang, U. Stimming, and A. A. Lee, "Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning," Nat. Commun., vol. 11, no. 1, p. 1706, Apr. 2020, doi: 10.1038/s41467-020-15235-7.

[10] S. Kohli, G. T. Godwin, and S. Urolagin, "Sales Prediction Using Linear and KNN Regression," in Advances in Machine Learning and Computational Intelligence, S. Patnaik, X.-S. Yang, and I. K. Sethi, Eds., in Algorithms for Intelligent Systems. Singapore: Springer Singapore, 2021, pp. 321–329. doi: 10.1007/978-981-15-5243-4_29.

[11] M. Kavitha, G. Gnaneswar, R. Dinesh, Y. R. Sai, and R. S. Suraj, "Heart Disease Prediction using Hybrid machine Learning Model," in 2021 6th International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India: IEEE, Jan. 2021, pp. 1329–1333. doi: 10.1109/ICICT50816.2021.9358597.

[12] N. Kholodna, V. Vysotska, and S. Albota, "A Machine Learning Model for Automatic Emotion Detection from Speech.," in MoMLeT+ DS, 2021, pp. 699–713.

[13] M. Hasan, E. Rundensteiner, and E. Agu, "Automatic emotion detection in text streams by analyzing Twitter data," Int. J. Data Sci. Anal., vol. 7, no. 1, pp. 35–51, Feb. 2019, doi: 10.1007/s41060-018-0096-z.

[14] H. AlBalooshi, S. Rahmanian, and R. Venkatesh Kumar, "EmotionX-SmartDubai_NLP: Detecting User Emotions In Social Media Text," in Proceedings of the Sixth International Workshop on Natural Language Processing for Social Media, Melbourne, Australia: Association for Computational Linguistics, 2018, pp. 45–49. doi: 10.18653/v1/W18-3508.

[15] H. Mahajan and S. Banerjee, "A machine learning framework for guided wave-based damage detection of rail head using surface-bonded piezo-electric wafer transducers," Mach. Learn. Appl., vol. 7, p. 100216, Mar. 2022, doi: 10.1016/j.mlwa.2021.100216.

[16] J. Lu et al., "SCueU-Net: Efficient Damage Detection Method for Railway Rail," IEEE Access, vol. 8, pp. 125109–125120, 2020, doi: 10.1109/ACCESS.2020.3007603.

[17] S. Moeinizade, H. Pham, Y. Han, A. Dobbels, and G. Hu, "An applied deep learning approach for estimating soybean relative maturity from UAV imagery to aid plant breeding decisions," Mach. Learn. Appl., vol. 7, p. 100233, Mar. 2022, doi: 10.1016/j.mlwa.2021.100233.

[18] D. Berrar, P. Lopes, and W. Dubitzky, "Incorporating domain knowledge in machine learning for soccer outcome prediction," Mach. Learn., vol. 108, no. 1, pp. 97–126, Jan. 2019, doi: 10.1007/s10994-018-5747-8.

[19] F. Haselbeck, J. Killinger, K. Menrad, T. Hannus, and D. G. Grimm, "Machine Learning Outperforms Classical Forecasting on Horticultural Sales Predictions," Mach. Learn. Appl., vol. 7, p. 100239, Mar. 2022, doi: 10.1016/j.mlwa.2021.100239.

[20] M. Kumar, A. Kumar, and V. S. Palaparthy, "Soil Sensors-Based Prediction System for Plant Diseases Using Exploratory Data Analysis and Machine Learning," IEEE Sens. J., vol. 21, no. 16, pp. 17455–17468, Aug. 2021, doi: 10.1109/JSEN.2020.3046295.

[21] K. P. Panigrahi, H. Das, A. K. Sahoo, and S. C. Moharana, "Maize Leaf Disease Detection and Classification Using Machine Learning Algorithms," in Progress in Computing, Analytics and Networking, H. Das, P. K. Pattnaik, S. S. Rautaray, and K.-C. Li, Eds., in Advances in Intelligent Systems and Computing, vol. 1119. Singapore: Springer Singapore, 2020, pp. 659–669. doi: 10.1007/978-981-15-2414-1_66.

[22] K. Ahmed, T. R. Shahidi, S. Md. Irfanul Alam, and S. Momen, "Rice Leaf Disease Detection Using Machine Learning Techniques," in 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI), Dhaka, Bangladesh: IEEE, Dec. 2019, pp. 1–5. doi: 10.1109/STI47673.2019.9068096.

[23] W. Khan, M. A. Ghazanfar, M. A. Azam, A. Karami, K. H. Alyoubi, and A. S. Alfakeeh, "Stock market prediction using machine learning classifiers and social media, news," J. Ambient Intell. Humaniz. Comput., vol. 13, no. 7, pp. 3433–3456, Jul. 2022, doi: 10.1007/s12652-020-01839-w.

[24] W. Ahmad et al., "Towards Short Term Electricity Load Forecasting Using Improved Support Vector Machine and Extreme Learning Machine," Energies, vol. 13, no. 11, p. 2907, Jun. 2020, doi: 10.3390/en13112907.

[25] S. A. Suha and M. N. Islam, "An extended machine learning technique for polycystic ovary syndrome detection using ovary ultrasound image," Sci. Rep., vol. 12, no. 1, p. 17123, Oct. 2022, doi: 10.1038/s41598-022-21724-0.

[26] A. L. Georgescu, J. C. Koehler, J. Weiske, K. Vogeley, N. Koutsouleris, and C. Falter-Wagner, "Machine Learning to Study Social Interaction Difficulties in ASD," Front. Robot. AI, vol. 6, p. 132, Nov. 2019, doi: 10.3389/frobt.2019.00132.

[27] M. N. Parikh, H. Li, and L. He, "Enhancing Diagnosis of Autism With Optimized Machine Learning Models and Personal Characteristic Data," Front. Comput. Neurosci., vol. 13, p. 9, Feb. 2019, doi: 10.3389/fncom.2019.00009.

[28] L. Xu, X. Geng, X. He, J. Li, and J. Yu, "Prediction in Autism by Deep Learning Short-Time Spontaneous Hemodynamic Fluctuations," Front. Neurosci., vol. 13, p. 1120, Nov. 2019, doi: 10.3389/fnins.2019.01120.

[29] A. A. Abdullah, S. Rijal, and S. R. Dash, "Evaluation on Machine Learning Algorithms for Classification of Autism Spectrum Disorder (ASD)," J. Phys. Conf. Ser., vol. 1372, no. 1, p. 012052, Nov. 2019, doi: 10.1088/1742-6596/1372/1/012052.

[30] R. O. Bahado-Singh et al., "Artificial intelligence analysis of newborn leucocyte epigenomic markers for the prediction of autism," Brain Res., vol. 1724, p. 146457, Dec. 2019, doi: 10.1016/j.brainres.2019.146457.

[31] C. Radhika and D. N. Priya, "Advance Hybrid RF-GBC-RFE Wrapper-Based Feature Selection Techniques for Prediction of Autistic Disorder," J. Algebr. Stat., vol. 13, no. 1, 2022.

[32] K. Vakadkar, D. Purkayastha, and D. Krishnan, "Detection of Autism Spectrum Disorder in Children Using Machine Learning Techniques," SN Comput. Sci., vol. 2, no. 5, p. 386, Sep. 2021, doi: 10.1007/s42979-021-00776-5.

[33] A. Riaz, M. Asad, E. Alonso, and G. Slabaugh, "DeepFMRI: End-to-end deep learning for functional connectivity and classification of ADHD using fMRI," J. Neurosci. Methods, vol. 335, p. 108506, 2020.