

Research Article

Predicting Risk through Artificial Intelligence Based on Machine Learning Algorithms: A Case of Pakistani Nonfinancial Firms

Shamsa Khalid,¹ Muhammad Anees Khan ,² M.S. Mazliham,³
Muhammad Mansoor Alam ,^{3,4,5,6} Nida Aman,² Muhammad Tanvir Taj,² Rija Zaka,²
and Muhammad Jehangir⁷

¹Master in Finance Scholar, Management Studies Department Bahria Business School, Islamabad, Pakistan

²Department of Management Studies Bahria University, Islamabad, Pakistan

³Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia

⁴Faculty of Computer and Information, Multimedia University, Cyberjaya, Malaysia

⁵Malaysian Institute of Information Technology, Universiti of Kuala Lumpur-Malaysia, Kuala Lumpur, Malaysia

⁶Faculty of Computing Riphah International University, Islamabad, Pakistan

⁷Abdul Wali Khan University Mardan, Mardan, Pakistan

Correspondence should be addressed to Muhammad Anees Khan; manees.buic@bahria.edu.pk

Received 5 November 2021; Revised 17 February 2022; Accepted 3 March 2022; Published 7 June 2022

Academic Editor: Atila Bueno

Copyright © 2022 Shamsa Khalid et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

AI (artificial intelligence) is a significant technological advancement that has everyone buzzing about its incredible potential. The current research study evaluates the influence of supervised artificial intelligence techniques, i.e., machine learning techniques on the nonfinancial firms of Pakistan and focuses on the practical application of AI techniques for the accurate prediction of corporate risks which in turn will lead to the automation of corporate risk management. So, in this study, we used financial ratios for accurate risk assessment and for the automation of corporate risk management by developing machine learning algorithms using techniques, namely, random forest, decision tree, naïve Bayes, and KNN. A secondary data collection technique will be used. For this purpose, we collected annual data of nonfinancial companies in Pakistan for the period ranging from 2006 to 2020, and the data are analyzed and tested through Python software. Our results prove that AI techniques can accurately predict risk with minimum error values, and among all the techniques used, the random forest technique outperforms as compared to the rest of the techniques.

1. Introduction

Artificial intelligence is a major technological breakthrough that is the need of the day for solving the different risk assessments and predictions issues. Profound learning and machine learning (ML) come under the heading of the artificial intelligence. AI is described as machines' ability to make intellectual human-like decisions and continue to improve. Machine learning entails the development of models, primarily statistical models that can be built and provide predictive results. Initially, AI incorporation in developing software was limited to big corporations with the financial means to employ highly skilled personnel. AI

frameworks with high levels of abstraction have been built over time, and an intelligent system can be created with just a few lines of code in any programming language of choice. In a comparison to conventional techniques that are quickly becoming obsolete, AI techniques may offer significant benefits to the world of finance, by automating certain tasks and boosting analytical capability. AI is a critical part of modern finance because it makes it cheaper, quicker, bigger, more available, profitable, and competitive in a variety of ways.

FERMA published the first thinking paper on the implications of artificial intelligence (AI) for risk management in November 2019, with the aim of guiding risk managers

from a basic understanding to designing their own artificial intelligence implementation plan. Moreover, another paper presented in which they focused on using techniques of artificial intelligence and BD in the integration of open libraries (BigData) and neural network bases, which play an important function in the automation of corporate risk management. The proposed solution, that is, distributed analysis system, is developed by using deep learning algorithms. Similarly, Aziz and Dowling conducted a research study in France and explored how risk management can be transformed by using machine learning and AI. Neither of such studies has been conducted in the nonfinancial sector of Pakistan where corporate risk management could have been automated using AI techniques such as random forest, ANN, deep learning, and so on. The problem statement that will be addressed in this study is how AI techniques can automate corporate risk management, and that of the AI techniques is the best fit for accurately determining how much risk companies will face.

1.1. Basic Understanding of Artificial Intelligence

1.1.1. Why Artificial Intelligence (AI)? Artificial intelligence develops a system that is intellectual and self-contained. Basically, artificial intelligence techniques are used for building prediction models in the form of algorithms with the highest possible accuracy. AI is rapidly evolving in a variety of fields and has a plethora of features. The technology has the potential to be used in a variety of industries and sectors. Artificial intelligence is applied in finance to recognize and track financial and banking activities such as unusual debit card use and large deposits of account, both of which help a fraud unit of the bank. The application of AI makes trading easier and more effective. It can be accomplished by making it easier to estimate production, demand, and stock price [1]. Machine learning comes under the heading of artificial intelligence that allows a system to adopt and develop its own understanding without having to program it explicitly. AI works in two ways: one is data-driven, and the other is symbolic. For the data-based side called ML, a large amount of data is needed to be fed into the machine before it is capable of learning. Machines are capable of learning in a much wider range of dimensions. Machines can deduce patterns from large amounts of high-dimensional data. Once these models are mastered, they will generate predictions that humans are unable to match [2]. Reasoning, information representation, NLP, scheduling, deep learning, interpretation, and the ability to transfer and control them are all domains of AI study. One of AI's long-term goals is to achieve general intelligence. The procedures that are needed to achieve the objective comprise traditional symbolic AI, artificial intelligence, and statistical methods. AI research is ongoing and evolving in the present day. China is expected to overtake the United States in the upcoming four years as the leading hub of AI, having taken over the second spot from the United States in 2004 and rapidly approaching Europe's main spot. Europe is regarded as the most diverse and largest region in the context of global

cooperation in AI studies. After the United States and China, India comes the third number to be the largest AI research country. AI is important and developed to the extent that a Japanese venture capital company made a name for itself by appointing the first AI board member to predict industry dynamics faster than humans [3].

1.1.2. What Is Machine Learning (ML)? Machine learning is basically the artificial intelligence's subfield that involves the automatic detection of patterns in data. It was originally described as a program that learns to accomplish a task automatically from experience (i.e., data) rather than being programmed explicitly. Machine learning algorithms have the capability to evolve through training, and they can process vast volumes of data and extract meaningful information using a variety of programming techniques. In this way, they can learn from the data they are given and improve on previous iterations [4]. In the finance industry, machine learning is being used to improve a variety of functions, namely fraud detection, payment processing, and regulation [5].

1.1.3. Types of Machine Learning (ML) Techniques. Machine learning systems are categorized based on how much and what kind of supervision they receive during training. Machine learning is broadly divided into the following three categories:

(1) *Supervised Learning.* Based on a set of training data, supervised learning develops a function that translates inputs to outputs. In the training set, the algorithm infers a function that connects each set of inputs to the predicted, or labeled, output [5]. The purpose of supervised learning is to predict a known outcome. To train a model, a data set with features (variables) and labels (outcome or class of interest) is used. The technique generates a function that maps features to labels and then utilizes it to predict the labels of unlabeled data. The accuracy of supervised learning models in predicting outcomes across one or more sets of data not included in the development process is typically measured. The most important supervised learning algorithms are random forest, decision tree, K-nearest neighbors, support vector machine (linear), support vector machine (RBF), naïve Bayes, polynomial regression, and artificial neural networks. Some of these are also used in this research study for the prediction of risk.

(2) *Unsupervised Learning.* In unsupervised learning, the training data are not labeled. The system attempts to learn without the assistance of a teacher. Unsupervised learning uncovers hidden patterns in unlabeled data and makes conclusions from it. Unsupervised learning feeds data into models but does not describe a set of predicted outcomes; the results are unlabeled [5]. Unsupervised learning is not anticipation of a specific outcome. Instead, the program looks for patterns or groups of data to identify.

(3) *Reinforcement Learning*. Reinforcement learning is a relatively newer class of learning that combines supervised and unsupervised learning. It is a completely distinct beast. The algorithm in reinforcement learning maximizes accuracy through trial and error. The model is shaped through feedback from the training set's outcomes of real and simulated decisions. In this context, the learning system is referred to as an agent since it can monitor the environment, select, and conduct actions and receive rewards (or penalties in the form of negative rewards). It must then figure out for itself what the ideal technique, known as a policy, is for maximizing reward over time. In each case, a policy specifies what action the agent should take.

1.1.4. Risks in Corporate Sector. Although there are several risk types, the most evident ones in the corporate sector that we are including in our study as well are financial, operational, strategic, and reputational risks. In reality, studies show that financial risks account for just around 10% of major market capitalization declines, while operational risks account for about 30%; the remaining 60% of declines are due to strategic risks, despite the fact that strategy ranks third in risk-prioritization exercises [6]. "Financial risk is the prospect of losing money on an investment or business plan." This risk can cause those who are exposed to it to lose their money and capital. Popular financial risk types include "credit risk, liquidity risk, asset-backed risk, foreign investment risk, equity risk, and currency risk." The probability of loss because of ineffective or procedures' failure, structures, or practices; employee mistakes; system failures; fraud or other criminal activities; or any incident that destabilizes business processes is referred to as operational risk. [7]. "A reputational risk is a hazard or danger to a company's or entity's good name or status. This type of risk may arise in multiple ways either directly due to the company's activities, or indirectly due to the actions of an employee or employees, or tangentially via other peripheral parties such as joint venture partners or suppliers." "Strategic risks are those that occur as a result of fundamental decisions made by directors on an organization's objectives." Different financial risk ratios are used by investors to evaluate a prospect of a company [8]. Our main objective is to evaluate the influence and practical application of artificial intelligence technologies on the automation of corporate risk management in Pakistani nonfinancial firms. In addition to evaluating the influence of AI, we will also develop an algorithm using python for accurate prediction of financial, operational, strategic, and reputational risks in the corporate sector. Furthermore, we will compare the accuracy level of different artificial intelligence techniques to determine which technique will be the best fit for risk assessment.

This research study is organized as follows: Section 2 examines related research. Section 3 describes the proposed methodology and model for risk prediction utilizing different machine learning algorithms and techniques.

Section 4 presents the analysis and results, while Section 5 analyses the managerial, social, and methodological implications of this research study. Finally, in Section 6, the investigation ends.

2. Related Research

AI must be able to cope with the idea of risk, which is the central domain of finance, in order to be useful to the financial system. Credit risk can be calculated using a variety of machine learning approaches that can extract nonlinear relationships between financial data on the balance sheets. Models are chosen in a typical data science life cycle to maximize predicted accuracy. The decisions are improved by using a posteriori explanations algorithm and model selection based on their accuracy prediction [9].

Financial risk early warning is established by relying on information technologies, fuzzy mathematical models, and other methods of analysis for developing a suitable financial risk early-warning index system, which is primarily based on internal control information of enterprises, external environmental data, industry data, and financial statement data. The financial position of the firm is examined; probability of financial hazards in the production is estimated; and the operation process based on the dynamic changes of a series of indicators is also examined. In a nutshell, financial risk early warning refers to the process of detecting, monitoring, and controlling a company's financial hazards [10]. Taking on risk is an inevitable result of making investments, and it is therefore necessary for economic activity. Too much risk, on the other hand, may result in unsustainable losses and even systemic crises. Unfortunately, the risk is a latent variable that must be derived from observable results rather than a single definition. As a result, any risk assessment is subjective based on a statistical model that must be assumed. There are a plethora of "risk meters" to choose from, many of which have very different results, making it difficult to distinguish between them [11].

AI use is often considered in equation with automation, in which computers are replacing human beings in occupations and decision-making, but artificial intelligence is applied to augment human interaction, such as risk or priority tracking and monitoring in recruiting, partially self-driving cars with human override, suggested customer service scripts, and audits, fraud detection, and judicial sentencing [12]. "The emergence of AI has radically altered trading, financial research, risk analysis, wealth management, investment banking, and other areas of the financial industry, resulting in profit enhancement and social benefits." AI has lowered capital costs for companies and entrepreneurs, widened the types of financial services available to a larger and more diverse population of investors, and made it easier for customers to bank and invest, but it also comes with significant risks [13].

Since the 1970s, the use of computer resources for decision making comprising risk was extensively researched in the field of information technology as decision support systems [14]. In 2012, Otim et al. published a study that focused on assessing the benefit and risks of IT investments.

Such investments entail a diverse range of stakeholders, necessitating the consideration of organizational politics [15]. Multiple stakeholders are involved in industrial decision-making, as are multiple parameters, which are influenced in part by the presence of these multiple stakeholders. For risk analysis in manufacturing multiple criteria, risk evaluation technique can also be applied [16]. Moreover, many other frameworks are also proposed for assessing risk and other problems early in the product development process. Financial markets are nonlinear, dynamic structures with subtleties and interactions that are difficult for humans to grasp. Due to this reason, artificial neuro networks have been widely used in this field. ANNs may be used to forecast currency markets, bank liquidity, inflation, and a variety of other financial needs. “Financial modelling, forecasting investor behavior, financial assessment, credit approval, asset portfolio management, pricing initial public offerings, and evaluating optimal capital structure are just a few of the corporate finance applications that can benefit from ANN technology” [17]. Neural networks are the tools of artificial intelligence that have proven to be extremely effective at detecting patterns in complex data structures, especially those with nonlinear relationships [14].

AI can gradually provide reliable real-time information on all forms of risks that the company is taking. Real-time advice will become more prevalent as data organization becomes more oriented toward AI use. Preemptive warning of risks is the next step after real-time awareness of risks being taken. An AI-driven risk management framework enables businesses to reliably predict firm risks, such as economy, credit, and operational risk, in advance. Machine learning techniques have this capability in a way that conventional statistical techniques could never hope to match [18]. Financial big data includes relevant financial details such as “interbank liquidity” and “global capital flow” that can be used to perform tasks such as risk early warning and risk discerning in order to react to risk through financial regulation [19]. “Big data has been used to analyze the interrelationships between risk source and risk diversification based on financial big data for systemic financial risk. The application of real data in financial systematic risk is one of the main challenges as financial data is kept under strict control” [20].

Artificial intelligence can help organizations at different stages of the risk management process, including determining risk exposure, evaluating, estimating, and analyzing its impact [21]. It may also aid in the selection of an effective risk reduction plan and the identification of instruments that promote risk shifting or trading. Thus, the use of AI strategies for organizational risk management has extended to new areas such as reviewing detailed paperwork and conducting routine procedures, which began with attempting to avoid external losses such as credit card fraud as well as detection of money laundering that requires analysis of large data sets.

In summary, this work adds to and extends the existing literature in various ways. First, it extends the application of artificial intelligence technologies in the nonfinancial sector of Pakistan by providing a detailed overview of the techniques

used for accurate risk prediction. It provides the basic information regarding the use of different techniques such as random forest, decision tree, naïve Bayes, and KNN. Second, it develops a framework along with the machine learning algorithms to help the practitioners and industrialists make use of the best-fit technique of AI offering the highest accuracy level and least error values for the first time in the nonfinancial sector of Pakistan. Finally, we examine how these technologies can enhance the accuracy level of risk assessment in the corporate sector by testing the data using python software.

3. Methodology

This paper aims to predict risk by offering the highest accuracy level by applying artificial intelligence based on machine learning algorithms that will lead to the automation of corporate risk management. For the accurate prediction of various types of risks in the corporate sector through artificial intelligence technologies based on machine learning algorithms, a secondary data collection technique is used, and quantitative data is gathered for this research study. Stratified random sampling is utilized in this research study. The total number of companies that are registered at the Pakistan stock exchange (PSX) is 540: 443 companies are of nonfinancial sector, and 97 companies are of financial sector. Out of these 443 companies, the active nonfinancial companies whose data was available was 330. Companies were then selected using a stratified sampling technique for this research; the sample of 200 firms in nonfinancial sector of Pakistan, ranging from 2006 to 2020, have been selected for this research to determine the accurate prediction of the risk of companies of nonfinancial registered at the Pakistan stock exchange (PSX). A total of 200 nonfinancial companies that have a high market capitalization value are selected as samples. For the prediction of corporate risks, that is, financial risk, reputational risk, operational risk, and strategic risk, and for developing the algorithms for different machine learning techniques, the variables used are debt-to-equity ratio, debt-to-capital ratio, degree-of-combined leverage, interest coverage ratio, debt-to-asset ratio, and equity ratio. Python version 3.9.2 is used in this study to test the data for risk prediction of nonfinancial firms listed on the Pakistan stock exchange. Different machine learning algorithms that are suited for this scenario were used to make the prediction.

After data have been collected, descriptive statistics and visualizations are used to assess their properties. Preparation of data set for analysis purposes so that it may be used as input for the algorithm is one of the most difficult processes. Structure of data, frequency of sample, missing values, and omics inclusion all should be considered, possibly needing varied knowledge levels. In this paper, the data set was first preprocessed to identify financial ratios that are used to predict various types of risk on the corporate side. The data set was partitioned into two subsets after the initial data preprocessing step: a training set (from 2006 to 2018) for training the models and a testing set (from 2019 to 2020) for testing the models. Two thousand and six hundred rows are utilized for training, and 400 rows are used for testing from the current data set. The next step was to identify which

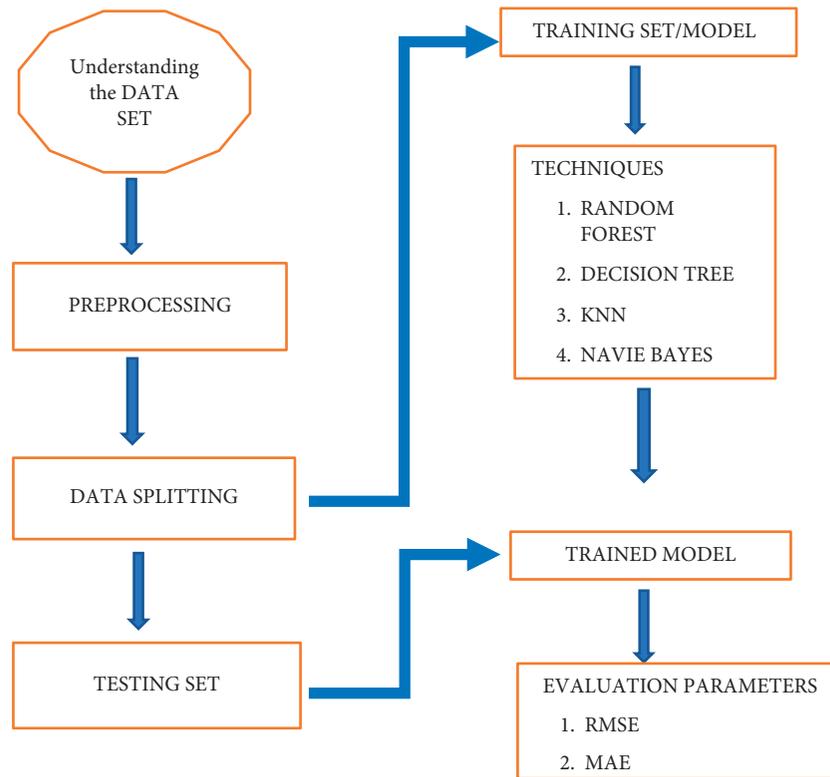


FIGURE 1: Framework for ML studies in the context of risk prediction.

models will be used and how data will be partitioned. This research used learning models such as random forest, decision tree, KNN, and naïve Bayes. These models were developed to estimate risk based on financial ratios. The learning models were then assessed using critical metrics including RMSE and MAE, and the findings were presented. Figure 1 depicts the proposed approach for the whole process of conducting the research.

3.1. Techniques Used. Over the past few years, there has been a proliferation of machine learning approaches and an increased surge in implications of these approaches in finance. These approaches have been employed in the optimization of portfolios, risk modeling, trend and sentiment analysis of news, and a variety of other use cases that help investment management.

Classification and regression are the key research areas in supervised learning, and this method is frequently employed in constructing prediction models. In contrast to traditional regression, regression machine learning employs regression algorithms that allow many variables to be utilized as independent variables before being automatically removed if they lack explanatory power. Because the data scientist has access to a significant amount of data, this is a vital feature. It also cuts down on the amount of reasoning required to find appropriate independent variables [18].

3.1.1. Random Forest (RF). Random forest is a supervised learning technique that constructs many decision trees during the training period and outputs the class that is the

mode of the classes or the mean prediction of the individual trees. It is used for classification, regression, and other applications. A decision tree is a tree structure that is used to make decisions (which can be a binary tree or a nonbinary tree). Every one of its nonleaf nodes represents a feature test, with each branch reflecting the feature attribute's output throughout a range of values and each leaf node holding a category. Moreover, Figure 2 has been shown diagrammatically that shows the bunch of random forest models. The classifier RF's goal is to combine many binary DTs created by generating several bootstrap instances from the learning sample L and randomly picking a subgroup of illustrated predictors X at each node. In their study, Uddin et al. used an upgraded version of RF produced by Salford Systems, with some special additions and modifications. By selecting the target variable and dependent attribute, the RF approach evaluates business intelligence (BI) and other domains, such as data sets, which is the main topic of this study. The advantage of RF is that it can automatically choose predictors from a large number of possibilities; for example, RF methods can determine the best predictor from thousands of variables [22].

3.1.2. Decision Tree. A decision tree is a supervised machine learning technique that can be used to solve problems in classification and regression. A decision tree is nothing more than a set of consecutive decisions that lead to a certain outcome. We can see the diagrammatic illustration of the decision tree model technique in Figure 3 that is, the supervised technique of machine learning, and used later in the

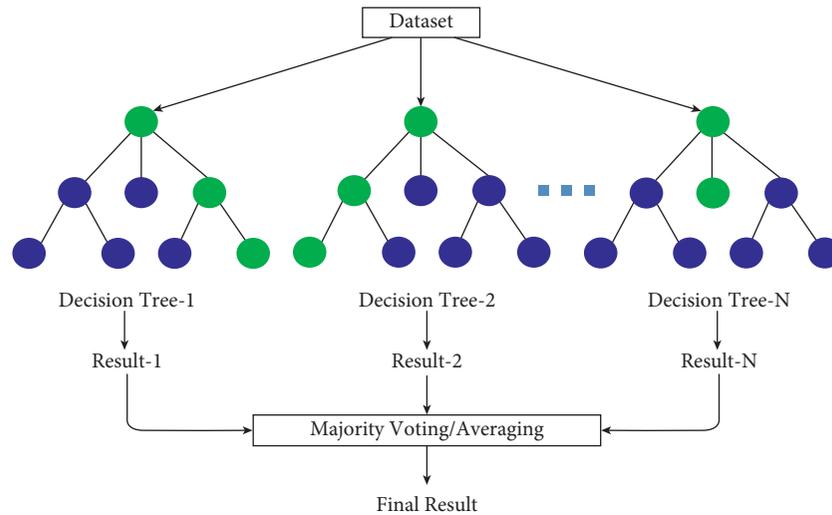


FIGURE 2: Random forest model (adapted from Analytics Vidhya, 2020).

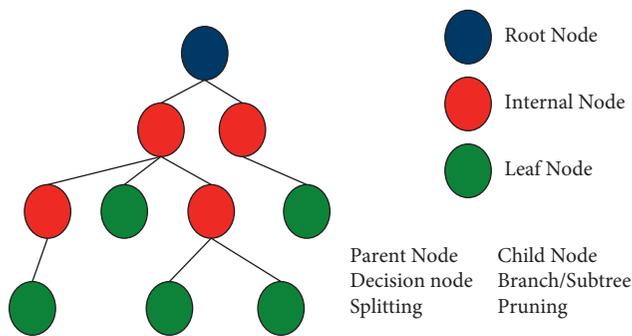


FIGURE 3: Illustration of the decision tree model.

methodology for prevention of the risk. The features/attributes and conditions can change depending on the data and the problem's complexity, but the core concept remains the same. A decision tree, thus, generates a sequence of judgments based on a set of features/attributes present in the data.

3.1.3. Naïve Bayes. The naïve Bayes classifier ignores probable input dependencies (correlations). That is, for a given class variable, all feature variables are conditionally independent of one another. After the class probability estimation, the final decision regarding a class is made [23]. Naïve Bayes classifiers, which are based on the well-known Bayes' probability theorem, are known for producing simplistic yet effective models, particularly in the domains of document classification and risk prediction. The Bayes theorem underpins the probabilistic model of naïve Bayes classifiers, and the adjective naïve refers to the assumption that the characteristics in a data set are mutually independent. Naïve Bayes classifiers are utilized in a variety of domains because they are generally robust, straightforward to implement, rapid, and accurate. Some instances include disease diagnosis and treatment decisions, classification of RNA sequences in taxonomic investigations, spam filtering in e-mail clients, and many others. The illustration of this technique is shown in Figure 4.

3.1.4. KNN. It is a nonparametric classifier that makes predictions without establishing a model by using a distance measure. The class labels and feature vectors of training samples are stored during the KNN training stage. For the testing stage, the distance between the new vector and all previously stored vectors is calculated, and the K closest samples are chosen. The class label is then assigned based on most of the k -nearest neighbors' classes [23].

“The K-NN method assumes that the new case/data and existing cases are similar and places the new case in the category that is most like the existing categories. The K-NN algorithm saves all existing data and classifies fresh data points according to their similarity. This means that new data can be quickly sorted into a well-defined category using the K-NN method. The K-NN approach can be used for both regression and classification, but it is more commonly utilized for classification tasks. The K-NN algorithm is a nonparametric algorithm, which means it makes no assumptions about the underlying data. It's also known as a lazy learner algorithm since it doesn't learn from the training set right away; instead, it saves the data set and executes an action on it when it comes to classify it. The KNN method simply stores the data set during the training phase, and when it receives new data, it classifies it into a category that is quite like the new data.” It is illustrated in Figure 5.

3.2. Advantages of the Classifiers Using an ML Technique. Each type of classifier used as an ML technique can be evaluated as having some advantages and disadvantages to be used as a future predictor; for example, ANN has the advantages of user friendly and easy to be used for implementation of the results. The big advantage of the ANN is to develop simple and user friendly algorithm based on machine learning, which will provide effective solution in comparison with traditional estimation techniques. In short, we can say that ANN is versatile and is used for different applications and problem solutions. It is a versatile classifier [24]. It can be

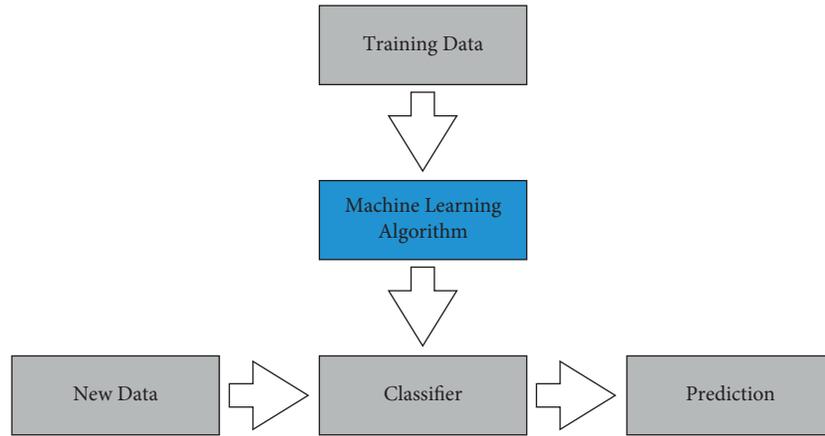


FIGURE 4: Naïve Bayes model (adapted from sebastianraschka.com).

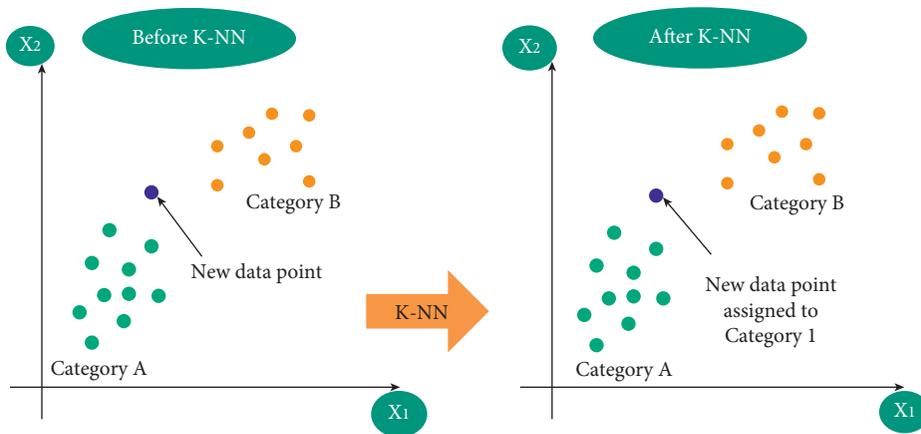


FIGURE 5: Illustration of KNN technique (adapted from JavaTpoint).

used for many applications and problems. However, besides the advantages, ANN also has some disadvantages; for example, it will get slower with a huge data that is ready for training [25]. In addition to the above advantages of the technique of the naïve Bayes classifier, we need small amount of training data, which will be converted into proper classification that can be easily implemented with a high recognition rate [26].

3.3. Model's Evaluation Parameters. In this research study, models' performance is evaluated based on two evaluation parameters, that is, root-mean-square error (RMSE) and mean absolute error (MAE). These measurements show us how accurate our forecasts are and how far off they are from the actual data.

3.3.1. Root-Mean-Square Error (RMSE). The difference between the values predicted by a model and their actual values is measured using the root-mean-square error (RMSE). The root-mean-square error (RMSE) is a widely used evaluation parameter, and it indicates how much error the system

makes in its predictions. The lower the value of RMSE, the better the model.

Mathematically, RMSE is computed by the following formula:

$$\text{MSE}(X, h) = \sqrt{\frac{1}{m} \sum_{i=1}^m (h(X^{(i)}) - Y^{(i)})^2}, \quad (1)$$

where m is the number of instances in the data set, $X^{(i)}$ is a vector containing all the feature values (excluding the label) of the i^{th} instance, $Y^{(i)}$ is its label (the desired output value for that instance), X is a matrix that contains all the feature values (excluding labels) of all instances in the data set, h is system's prediction function, also called a hypothesis, and $\text{RMSE}(X, h)$ is the function measured on the set of examples using hypothesis h .

3.3.2. Mean Absolute Error (MAE). The average of the absolute value of the errors is termed mean absolute error (MAE). The lower the value of MAE, the better the model.

Mathematically, MAE is computed by the following formula:

TABLE 1: Error measurement using KNN, naïve Bayes, decision tree, and random forest.

Technique	RMSE	MAE
Debt-to-asset ratio KNN	0.17256	0.08551
Debt-to-asset ratio naïve Bayes	0.25266	0.20181
Debt-to-asset ratio decision tree	0.09421	0.03592
Debt-to-asset ratio random forest 2	0.08171	0.03472
Debt-to-asset ratio random forest 3	0.09916	0.036
Debt-to-asset ratio random forest 5	0.08907	0.03084
Debt-to-asset ratio random forest 7	0.08312	0.03288
Debt-to-asset ratio random forest 10	0.08093	0.03134
Debt-to-asset ratio random forest 15	0.08365	0.03049

$$\text{MAE}(X, h) = \frac{1}{m} \sum_{i=1}^m |h(X^{(i)} - Y^{(i)})|. \quad (2)$$

4. Analysis and Results

In this study, we predicted risk using different machine learning techniques and models such as random forest, decision tree, KNN, and naïve Bayes, whose evaluation is based on RMSE and MAE. For processing data into the model, the data set of all the six features is divided into training set and testing set. A total of six variables (financial ratios) were used for the accurate prediction of risk that are debt-to-capital ratio, debt-to-equity ratio, interest coverage ratio, degree of combined leverage, debt-to-asset ratio, and equity ratio. Out of these six variables, debt-to-asset ratio and equity ratio are our target features whose prediction is done based on the remaining four variables. Based on the prediction results, we will select those models whose error rate will be comparatively lesser than the other model. Basically, the comparison criteria for which the technique predicts risk more accurately is determined by the error value. The lesser the value of error from both the evaluation parameters, the better the technique.

4.1. Analysis of First Target Variable: Debt-to-Asset Ratio.

In Table 1, the results of four techniques of machine learning are presented in which error measurement is done based on RMSE and MAE for the variable debt-to-asset ratio. Here, we will analyze which one of these four techniques is the best and most accurate predictor of risk with less error value.

The result obtained from the naïve Bayes technique shows the value of 0.25266 for RMSE and the value of 0.20181 for MAE. Likewise, the results of RMSE and MAE obtained from the KNN technique show lower values as compared to the naïve Bayes technique. The RMSE value obtained from KNN is 0.17256 that is lesser than the RMSE value of naïve Bayes (i.e., 0.25266). Similarly, the MAE value obtained by applying the KNN technique is 0.08551 that is also less as compared to MAE values obtained from naïve Bayes (i.e., 0.20181).

In Table 1, results of decision tree (individual) and random forests with 2, 3, 5, 7, 10, and 15 decision trees have also been shown that are applied to our data set for accurate prediction of risk using RMSE and MAE as evaluation matrices. We can see

TABLE 2: Comparison of order of all the techniques used.

Risk prediction using RMSE parameter: Random forest > decision tree > KNN > naïve Bayes
Risk prediction using MAE parameter: Random forest > decision tree > KNN > naïve Bayes

TABLE 3: Error measurement using KNN, SVM (linear and RBF), and naïve Bayes.

Technique	RMSE	MAE
Equity ratio KNN	0.13878	0.03791
Equity ratio naïve Bayes	0.30508	0.21779
Equity ratio decision tree	0.05684	0.02149
Equity ratio random forest 2	0.05204	0.02073
Equity ratio random forest 3	0.06619	0.02212
Equity ratio random forest 5	0.06825	0.02299
Equity ratio random forest 7	0.05338	0.0208
Equity ratio random forest 10	0.05194	0.02019
Equity ratio random forest 15	0.0638	0.01825

that the decision tree model has RMSE and MAE values of 0.09421 and 0.03592, respectively. But the random forest technique with a different number of decision trees gives relatively lower values of RMSE and MAE than the single decision tree model. If we make predictions based on the RMSE parameter, random forest technique with 10 decision trees gives the lowest value of RMSE, that is, 0.08093 as compared to single decision tree and random forest with 2, 3, 5, 7, and 15 decision trees. If we make predictions based on MAE parameter, random forest technique with 15 decision trees gives the lowest value of MAE, that is, 0.03049 as compared to the single decision tree and random forest with 2, 3, 5, 7, and 10 decision trees. Therefore, random forest with a greater number of decision trees gives more accurate results, and the reason of this technique to perform better than single decision tree is that it has power of numerous decision trees combined in it and it does not rely on a single decision tree's feature relevance. The meaning of the value bold in Table 1 is, AS RMSE and MAE is performance evaluation methods that decides the best outperform technique in the existing techniques used in the research paper based on least variation like shown in the quantitative bold figures which shows the least variation in the existing one.

4.1.1. Result of Analysis of the First Target Variable. We can say collectively that for both evaluation parameters, random forest technique outperforms with least values of RMSE and MAE. So, overall, we can say that from the above-mentioned techniques of table, random forest technique outperforms the rest of the three techniques, that is, decision tree, KNN, and naïve Bayes.

Finally, the order of accuracy of risk prediction using equity ratio variable and different machine learning techniques is shown in Table 2.

4.2. Analysis of the Second Target Variable: Equity Ratio.

In Table 3, the results of four techniques of machine learning are presented in which error measurement is done based on

TABLE 4: Comparison of the order of all the techniques used.

Risk prediction using RMSE parameter: Random forest > decision tree > KNN > naïve Bayes
Risk prediction using MAE parameter: Random forest > decision tree > KNN > naïve Bayes

RMSE and MAE for the variable equity ratio. Here, we will analyze which one of these four techniques is the best and most accurate predictor of risk with less error value.

The result obtained from the naïve Bayes technique shows the value of 0.30508 for RMSE and the value of 0.21779 for MAE. Likewise, the results of RMSE and MAE obtained from the KNN technique show lower values as compared to the naïve Bayes technique. The RMSE value obtained from KNN is 0.13878 that is lesser than the RMSE value of naïve Bayes (i.e., 0.30508). Similarly, the MAE value obtained by applying the KNN technique is 0.03791, which is also less as compared to MAE values obtained from naïve Bayes (i.e., 0.21779).

In Table 3, results of decision tree (individual) and random forests with 2, 3, 5, 7, 10, and 15 decision trees have also been shown that are applied to our data set for accurate prediction of risk using RMSE and MAE as evaluation matrices. We can see that the decision tree model has RMSE and MAE values of 0.05684 and 0.01825, respectively. But the random forest technique with a different number of decision trees gives relatively lower values of RMSE and MAE than the individual decision tree model. If we make predictions based on the RMSE parameter, the random forest technique with 10 decision trees gives the lowest value of RMSE, that is, 0.05194 as compared to the single decision tree and random forest with 2, 3, 5, 7, and 15 decision trees. If we make predictions based on the MAE parameter, the random forest technique with 15 decision trees gives the lowest value of MAE, that is, 0.01825 as compared to the single decision tree and random forest with 2, 3, 5, 7, and 10 decision trees. Therefore, the random forest with a greater number of decision trees gives more accurate results, and the reason for this technique to perform better than the single decision tree is that it has the power of numerous decision trees combined in it and it does not rely on a single decision tree's feature relevance. The bold values shown in Table 1 means that which technique is better based on RMSE and MAE values. In layman approach, RMSE and MAE decide that which AI technique will outperform among the techniques used in the study.

4.2.1. Result of Analysis of the Second Target Variable. We can say collectively that for both evaluation parameters, the random forest technique outperforms with the least values of RMSE and MAE. So, overall, we can say that from the above-mentioned techniques in Table 3, random forest technique outperforms the rest of the three techniques, that is, decision tree, KNN, and naïve Bayes.

Comparisons of machine learning techniques are shown in Table 4.

Diagrammatic representation of application of the proposed research



FIGURE 6: Application of proposed research in different fields (adapted from Moritz 2018).

5. Discussion

5.1. Benefit of This Research Study to Society

5.1.1. Managerial Implications. This study contributes both theoretically (i.e., to academia) and practically (i.e., to industry). For academia, this study will help provide a better understanding of artificial intelligence technologies in nonfinancial sector and the influence of AI techniques in risk prediction and automation of corporate risk management. It will provide the basic information regarding the use of different techniques such as random forest, decision tree, KNN, naïve Bayes, and so on and how these technologies can enhance the accuracy level of risk assessment in the corporate sector. It will also contribute as a source of information and indication for future research and bibliographic reviews. For industries, it will provide an insight into artificial intelligence technologies and practices that need to be adopted by organizations to better predict their financial, operational, strategic, and reputational risk. From the findings of this research, the firms can analyze the advantages of the incorporation of artificial intelligence technologies in risk prediction. This research will help them analyze the ease and importance of AI technologies for the automation of corporate risk, which will help them in the long run as well. These techniques will be particularly useful for promoting growth and diminishing human error by automating the companies' corporate risk. It will assist investors such as insurance companies, banks, venture capital firms, and mergers and acquisitions in making better credit decisions, simplifying tasks, and improving the company's financial efficiency. Companies can easily verify the credit of both potential partners and customers by automating the risk assessment process, which can help them determine if they are credit risks.

5.1.2. Social and Methodological Significance. In Pakistan, organizations such as PACRA have begun to employ artificial intelligence strategies to assist in the granting of credit scores to institutions. This algorithmic pattern will aid in the fast and accurate prediction of risk. The nonfinancial firms in Pakistan should strive to exploit the need for advanced statistical models to accurately estimate and reduce the risk that has grown more crucial than ever before as businesses grow larger and more complex. The application of proposed research (i.e., implication of AI) in different fields of finance is also shown in Figure 6. With previously applied statistical or simulation methods, precisely measuring the portfolio's exposure to the dynamic financial market is becoming increasingly difficult for major organizations with vast portfolios and sophisticated financial products. These AI approaches with the least error rate would serve as a beacon of light for initiating artificial intelligence in the field of finance in nonfinancial firms in Pakistan. They will benefit the firms in long run and can also be helpful for the nonfinancial firms by offering a full description of the latest technology, their prospective uses, and the likelihood of successful application and by predicting risk accurately. Our findings are valuable to both academics and practitioners with an interest in investment management, particularly quantitative investment.

6. Conclusion

This research study was aimed to investigate, evaluate, and develop machine learning algorithms that could accurately predict risk with minimum error for the nonfinancial firms listed on PSX. By offering a full description of the latest technology, their prospective uses, and the likelihood of successful application and by predicting risk accurately, these types of models could be employed in nonfinancial firms in their investment management, particularly quantitative investment.

6.1. Our Contributions. In this research study, we first gathered secondary data, and data were of quantitative nature. We utilized stratified random sampling and selected the sample of 200 firms in the nonfinancial sector of Pakistan, ranging from 2006 to 2020, for this research to determine the accurate prediction of the risk of nonfinancial companies registered at the Pakistan stock exchange (PSX). Companies having a high market capitalization value are selected as samples. For the prediction of corporate risks, that is, financial risk, reputational risk, operational risk, and strategic risk, and for developing the algorithms for different machine learning techniques, we used the variables that are debt-to-equity ratio, debt-to-capital ratio, degree of combined leverage, interest coverage ratio, debt-to-asset ratio, and equity ratio. Python version 3.9.2 is used in this study to test the data for risk prediction of nonfinancial firms listed on the Pakistan stock exchange. We used different machine learning algorithms that are suited for this scenario to make the prediction such as random forest, decision tree, K-nearest neighbor (KNN), and naïve Bayes, and we did the

performance analysis of risk prediction using various above-mentioned techniques of machine learning. The performance analysis of risk prediction using various ML approaches is analyzed by using the two evaluation parameters that are RMSE and MAE.

6.2. Findings and Conclusion. The results of the first target variable, that is, debt-to-asset ratio show that the RMSE and MAE values of random forest (i.e., 0.08093 and 0.03049) are comparatively low as compared to the other techniques such as decision tree, KNN, and naïve Bayes. So random forest technique outperforms compared to the rest of the techniques as it gives the least values of RMSE and MAE. The random forest technique is the best fit for accurate prediction of risk as it is giving the minimum error values, and the reason for this technique to perform better than other techniques is that it has the power of numerous decision trees combined in it and it does not rely on a single decision tree's feature relevance. So, for accurate prediction of risk of nonfinancial firms of Pakistan using debt-to-asset ratio as a target feature, random forest is the best fit and preferable technique.

The results of the second target variable, that is, equity ratio also show that the RMSE and MAE values of random forest (i.e., 0.05194 and 0.02019), are comparatively low as compared to the other techniques such as decision tree, KNN, and naïve Bayes. So the random forest technique outperforms compared to the rest of the techniques as it gives the least values of RMSE and MAE. The random forest technique is the best fit for accurate prediction of risk as it is giving the minimum error values, and the reason for this technique to perform better than other techniques is that it has the power of numerous decision trees combined in it and it does not rely on a single decision tree's feature relevance. So, for accurate prediction of risk of nonfinancial firms of Pakistan using debt-to-asset ratio as a target feature, random forest is the best fit and preferable technique.

Collectively, for both the target features, random forest is deemed to be the best performing technique among other supervised AI techniques, that is, decision tree, KNN, and naïve Bayes as it gives the highest accuracy rate by providing minimum error values of RMSE and MAE. This AI approach with the least error rate would serve as a beacon of light for initiating artificial intelligence in the field of finance in nonfinancial firms of Pakistan. It would help creditors determine credit risk and would help investors look for the accurate risk prediction before making an investment in companies.

6.3. Limitations of Study. This study is limited to Pakistani nonfinancial firms, and it did not cover nonfinancial firms of all South Asian countries due to the shortage of time. This research might be conducted on other similar capacity countries to authenticate its results further or to identify changes in results in relation to different markets and understand this variation in relation to them, to broaden its reach and verify its results. Moreover, two evaluation parameters are used in this study based on which prediction is

done. More evaluation parameters such as accuracy, F1 score, and confusion matrix can also be incorporated to further validate the results. The analysis, discussion, and comparisons of only ML algorithms were used in this current study.

6.4. Future Work. Future researchers can focus on the implications and limitations to improve the quality of research in the future. This study has been conducted on nonfinancial firms in Pakistan only. Future researchers can incorporate the data of other south Asian countries as well and can broaden this study of predicting risk in nonfinancial firms of other South Asian countries. The current study needs to be expanded and enriched with a more detailed analysis to work on the hypothesis that the random forest technique is the best predictor than decision tree, KNN, and naïve Bayes using other variables for risk prediction instead of financial ratios. More evaluation parameters such as accuracy, F1 score, and confusion matrix can also be incorporated to further validate the results. In future, we can use both ML approaches and econometric forecasting models for more precise results and can check whether the hybrid of both these methods outperforms or which of the individual method outperforms.

Data Availability

This study is the analysis of existing secondary data, which are openly available at locations of the DataStream database and company annual reports of the samples. Thus, we used the secondary data that are available on the DataStream database and company annual reports of the samples of the paper that are already cited in the methodology section of the paper. Moreover, the data are open and available to everyone from the DataStream database.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] "Artificial intelligence." Available: <https://www.investopedia.com/terms/a/artificial-intelligence-ai.asp>, 2019.
- [2] A. Takyar, "What is Artificial Intelligence? understand artificial intelligence in 5 minutes," 2018, [https://www.leewayhertz.com/what-is-artificial-intelligence-understand-ai-in-5-minutes/%23:%7E:text=Artificial%20Intelligence%20\(AI\)%20is%20a,reasoning%2C%20and%20self-correction](https://www.leewayhertz.com/what-is-artificial-intelligence-understand-ai-in-5-minutes/%23:%7E:text=Artificial%20Intelligence%20(AI)%20is%20a,reasoning%2C%20and%20self-correction).
- [3] C. Aguis, "Evolution of AI: past, present, future," 2019, <https://www.analyticsinsight.net/the-evolution-of-artificial-intelligence-past-present-future/>.
- [4] R. Shouval, J. A. Fein, B. Savani, M. Mohty, and A. Nagler, "Machine learning and artificial intelligence in haematology," *British Journal of Haematology*, vol. 192, no. 2, pp. 239–250, 2020.
- [5] S. Emerson, R. Kennedy, L. O'Shea, and J. O'Brien, "Trends and applications of machine learning in quantitative finance," in *Proceedings of the 8th International Conference on Economics and Finance Research (ICEFR 2019)*, WikiCFP, Lyon, France, June 2019.
- [6] L. McMenemy, "Strategies for corporate risk management," 2019, <https://insights.diligent.com/risk-management/strategies-corporate-risk-management>.
- [7] I. A. Moosa, *Operational Risk Management*, Palgrave Macmillan, New York, NY, USA, 2007.
- [8] "Investopedia," 2021, <https://www.investopedia.com/terms/f/financialrisk.asp>.
- [9] N. Bussmann, P. Giudici, D. Marinelli, and J. Papenbrock, "Explainable AI in fintech risk management," *Frontiers in Artificial Intelligence*, vol. 3, p. 26, 2020.
- [10] M. Zhai, "Risk prediction and response strategies in corporate financial management based on optimized BP neural network," *Complexity*, vol. 2021, Article ID 9973377, 10 pages, 2021.
- [11] J. Danielsson and C. Zhou, "Why risk is so hard to measure," De Nederlandsche Bank Research Paper No 494, 2016.
- [12] A. Agrawal, J. S. Gans, and A. Goldfarb, "Exploring the impact of artificial intelligence: Prediction versus judgment," *Information Economics and Policy*, vol. 47, pp. 1–6, 2019.
- [13] W. Magnuson, "Regulating fintech," *Vanderbilt Law Review*, vol. 71, p. 1167, 2018.
- [14] D. D. Wu, S. H. Chen, and D. L. Olson, "Business intelligence in risk management: Some recent progresses," *Information Sciences*, vol. 256, pp. 1–7, 2014.
- [15] R. K. Kozhikode and J. Li, "Political pluralism, public policies, and organizational choices: Banking branch expansion in India, 1948–2003," *Academy of Management Journal*, vol. 55, no. 2, pp. 339–359, 2012.
- [16] A. Silvestri, F. De Felice, and A. Petrillo, "Multi-criteria risk analysis to improve safety in manufacturing systems," *International Journal of Production Research*, vol. 50, no. 17, pp. 4806–4821, 2012.
- [17] A. Bahrammirzaee, "Artificial neural networks, expert system and hybrid intelligent systems," *Neural Computing & Applications*, vol. 19, no. 8, pp. 1165–1195, 2010.
- [18] S. Aziz and M. Dowling, "Machine learning and AI for risk management," in *Disrupting Finance* Palgrave Pivot, Cham, Switzerland, 2019.
- [19] M. Campbellverduyn, M. Goguen, and T. Porter, "The case of financial practices," *New Political Economy*, vol. 22, no. 2, pp. 219–236, 2017.
- [20] P. Cerchiello and P. Giudici, "Conditional graphical models for systemic risk estimation," *Expert Systems with Applications*, vol. 43, pp. 165–174, 2016.
- [21] A. Sanford and I. Moosa, "A Bayesian network approach," *Journal of the Operational Research Society*, vol. 66, no. 1, pp. 86–115, 2015.
- [22] M. S. Uddin, G. Chi, M. A. M. Al Janabi, and T. Habib, "Leveraging random forest in micro-enterprises credit risk modelling for accuracy and interpretability," *International Journal of Finance & Economics*, 2020.
- [23] N. Arora and P. D. Kaur, "A Bolasso based consistent feature selection enabled random forest classification algorithm: An application to credit risk assessment," *Applied Soft Computing*, vol. 86, Article ID 105936, 2020.
- [24] O. AlShorman, F. Alkahatni, M. Masadeh et al., "Sounds and acoustic emission-based early fault diagnosis of induction motor: A review study," *Advances in Mechanical Engineering*, vol. 13, no. 2, 2021.
- [25] A. Glowacz, R. Tadeusiewicz, S. Legutko et al., "Fault diagnosis of angle grinders and electric impact drills using acoustic signals," *Applied Acoustics*, vol. 179, Article ID 108070, 2021.
- [26] R. Kochan, O. Kochan, M. Chyrka, S. Jun, and P. Bykovyy, "Approaches of voltage divider development for metrology verification of ADC," in *Proceedings of the IEEE 7th International Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS)*, September 2013.