Using Ensemble Learning and Advanced Data Mining Techniques to Improve the Diagnosis of Chronic Kidney Disease

Muneer Majid[1]*, Yonis Gulzar[2]*, Shahnawaz Ayoub[3], Farhana Khan[4],
Faheem Ahmad Reegu[5], Mohammad Shuaib Mir[6], Wassim Jaziri[7], Arjumand Bano Soomro[8]
Glocal School of Science and Technology Glocal University, Delhi-Yamunotri Marg (State Highway 57),
Mirzapur Pole, Dist - Saharanpur, U.P. - 247121, India[1, 3, 4]
Department of Management Information Systems, College of Business Administration, King Faisal University,
Al-Ahsa 31982, Saudi Arabia[2, 6, 7, 8]
College of Computer Science and Information Technology, Jazan University, Jazan 45142, Saudi Arabia[5]
Department of Software Engineering, Faculty of Engineering and Technology, University of Sindh, Sindh, Pakistan[8]

***Abstract*—Kidney failure is a condition with far-reaching, potentially life-threatening consequences on the human body. Leveraging the power of machine learning and data mining, this research focuses on precise disease prediction to equip decision-makers with critical data-driven insights. The accuracy of classification systems hinges on the dataset's inherent characteristics, prompting the application of feature selection techniques to streamline algorithm models and optimize classification precision. Various classification methodologies, including K-Nearest Neighbor, J48, Artificial Neural Network (ANN), Naive Bayes, and Support Vector Machine, are employed to detect chronic renal disease. A predictive framework is devised, blending ensemble methods with feature selection strategies to forecast chronic kidney disease. Specifically, the predictive model for chronic kidney disease is meticulously constructed through the fusion of an information gain-based feature evaluator and a ranker search mechanism, fortified by the wrapper subset evaluator and the best first algorithm. J48, in tandem with the Info Gain Attribute Evaluator and ranker search system, exhibits a remarkable accuracy rate of 97.77%. The Artificial Neural Network (ANN), coupled with the Wrapper Subset Evaluator and the highly effective Best First search strategy, yields precise results at a rate of 97.78%. Similarly, the Naive Bayes model, when integrated with the Wrapper Subset Evaluator (WSE) and the Best First search engine, demonstrates exceptional performance, achieving an accuracy rate of 97%. Furthermore, the Support Vector Machine algorithm achieves a notable accuracy rate of 97.12% when utilizing the Info Gain Attribute Evaluator. The K-Nearest Neighbor Classifier, in conjunction with the Wrapper Subset Evaluator, emerges as the most accurate among the foundational classifiers, boasting an impressive prediction accuracy of 98%. A second model is introduced, incorporating five diverse classifiers operating through a voting mechanism to form an ensemble model. Investigative findings highlight the efficacy of the proposed ensemble model, which attains a precision rate of 98.85%, as compared to individual base classifiers. This research underscores the potential of combining feature selection and ensemble techniques to significantly enhance the precision and accuracy of chronic kidney disease prediction.**

Keywords*—Kidney; chronic kidney disease; support vector machine; k-nearest neighbors; artificial neural network; decision tree

I. Introduction

Large datasets can be mined for significant insights via data mining, which is seen as a necessary stage in the learning process. Its uses are widespread across various industries, including business, healthcare, education, science, government, etc. Data mining is frequently used in the medical sector to forecast diseases [1]. Developing efficient approaches for illness analysis, prediction, and detection is central to this critical area of research in the healthcare industry [2, 3]. Applications for data mining are frequently used in patient care systems, health information systems, and healthcare management and they significantly affect the analysis of disease survival [4, 5].

Data mining and classification approaches are essential for classifying, identifying, analyzing, and predicting disease datasets in the healthcare domain [6,7]. Medical datasets undergo comprehensive type, meticulous research, precise detection, and informed prediction through various classification methodologies. These encompass the sophisticated realm of Artificial Neural Networks (ANN), the discerning approach of K-Nearest Neighbors (KNN), the probabilistic insight of Naïve Bayes, the strategic branching of Decision Tress (J48, C4.5), the adept maneuvering of Support Vector Machines (SVM) etc.

The feature selection approach is crucial in data mining and machine learning as it plays a vital role in knowledge discovery, pattern identification, and statistical sciences [7]. Eliminating pointless attributes from the dataset is the primary goal of feature selection [8]. The refinement of classifier performance accuracy can be achieved by strategically removing specific details. Wrapper and filter approaches can classify feature selection techniques [9, 10].

Machine learning methods called ensembles combine predictions from various classifiers to increase prediction

*Corresponding Author
email: ygulzar@kfu.edu.sa, muneerbazaz@gmail.com

accuracy. An essential process for creating extremely accurate prediction models is the ensemble model. Ensemble models, such as random forest, bagging, boosting, stacking, and voting, are commonly employed in machine learning, data mining, and data science.

Chronic renal failure (CRF), another name for chronic kidney disease (CKD), is a severe and developing health problem worldwide. CKD is characterized by a slow decline in kidney function that impairs the functionality of the renal organs [4]. Due to the lack of apparent symptoms in the early stages, the start of renal failure may initially go unreported [6]. However, the effects of renal failure can seriously harm a person's general health and potentially have deadly implications.

According to the Global Burden of Disease Project, chronic kidney disease (CKD) has recently become a rapidly expanding global health concern. According to statistical statistics, the death rate for people with CKD increased significantly by 90% between 1990 and 2013 [7]. Presently, (CKD) stands as the thirteenth most prevalent contributor to mortality on a global scale. Moreover, CKD prominently ranks within the upper echelon of the five leading causes of death worldwide, as substantiated by findings from the esteemed research conducted by Kidney International [8]. According to the National Kidney Foundation, CKD affects around 10% of the world's population and causes millions yearly deaths [11]. The high mortality rates linked to CKD result from a lack of efficient treatments and a poor understanding of renal disease.

In developing nations, some kidney patients sometimes wait until their ailment has advanced before seeking treatment. This pattern helps explain why CKD is becoming more common [8]. However, detecting the illness at an early stage or during its start can decrease or even stop the occurrence of CKD. Early detection and management of kidney illness can be aided by diagnostic procedures such as blood testing, urine tests, kidney scans, and doctor consultations regarding additional symptoms of kidney disease.

By using feature selection strategies to lower the dimensionality of the features and ensemble models, which include various classifiers, this work focuses on evaluating the accuracy of the methods.

The remainder of this study is divided into the following sections: A literature review is presented in Section II, the techniques are described in Section III, the experimental test findings are shown in Section IV with a discussion that follows and the research is concluded in Section V.

II. Literature Review

Classification methodologies, the process of selecting relevant features, and the utilization of ensemble approaches stand as foundational pillars within the realms of machine learning and data analysis. Several research endeavors have been undertaken to employ these methodologies to classify disease datasets within the medical domain, and these endeavors have been extensively examined in the discipline. Numerous research studies have shown promising classification accuracy when using feature selection approaches, ensemble models, data mining and machine learning techniques to analyze medical datasets.

A study on the diagnosis of chronic renal illness using Support Vector Machines (SVM) and efficient feature selection techniques was carried out by Polat et al. [6]. To reduce dimensionality, they employed both wrapper and filter feature selection approaches. The study revealed that using Support Vector Machines (SVM) without feature selection led to an impressive accuracy of 97.75% in their analysis. By integrating SVM with a classifier subset evaluator and applying a greedy stepwise technique, the accuracy was enhanced to 98%. Similarly, utilizing Support Vector Machines (SVM) with a wrapper subset evaluator and leveraging a best-first search technique resulted in an elevated accuracy of 98.25%. Correspondingly, merging Support Vector Machines (SVM) with a classifier subset evaluator and applying a greedy stepwise method yielded an accuracy rate of 98.25%. Lastly, employing the best-first search strategy alongside SVM using the filter subset evaluator achieved the highest accuracy rate of 98.5%.

To predict cardiac disease, Bashir et al. [12] suggested an ensemble classifier based on a majority vote framework. The ensemble model was built using Nave Bayes, decision trees based on Gini Index and information Gain, memory-based learners, and Support Vector Machine (SVM), five heterogeneous classifiers. Their MV5 framework obtained an accuracy of 88.5% through trials utilizing stratified cross-validation, with a sensitivity of 86.96%, specificity of 90.83%, and an F-Measure of 88.85. The ensemble model's average accuracy increased compared to the individual base classifiers. The suggested method involved producing personal classifier judgments, successfully integrated to create the new combined model.

Bashir et al. [13] presented the HMV framework for medical decision support, which employs a multi-layer classifier for disease prediction. Their strategy focuses on assembling diverse classifiers into an ensemble model most effectively. Within their system’s framework, an array of discerning classifiers is harnessed, including but not limited to Naive Bayes, Linear Regression, Quadratic Discriminant Analysis, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), as well as Decision Trees meticulously constructed using both the Gini Index and the Information Gain criterion. Their HMV ensemble framework outperformed other prediction models in experiments, according to the results. The three components of the HMV framework are data collection and preprocessing, predicting unidentified class labels for test instances, and assessing the suggested HMV ensemble model. They attained the maximum disease categorization and prediction accuracy level using the HMV ensemble model on the chosen dataset.

Data mining techniques were used in a study by Khajehali et al. [14] to uncover parameters impacting pneumonia patients. They suggested a modeling strategy that included ensemble approaches for feature selection and classification with preprocessing, dimensionality reduction and unstructured data classification. They used the Bayesian Boosting method to build a model that identifies variables related to patient length of stay (LOS) in the hospital. The design of their

investigation included various preprocessing stages. SVM and ensemble approach like AdaBoost, Vote, Stacking, and Bayesian Boosting were used in the modelling process. Using a 10-fold cross-validation procedure, Bayesian Boosting, one of these classifier algorithms, was used for data analysis. Ten subsets were created from the dataset, with the training subset being chosen iteratively ten times. The training ensembles encompassed a comprehensive selection, incorporating nine available ten subsets. The findings demonstrated the efficiency of the Bayesian Boosting ensemble technique in forecasting pneumonia disease and anticipating length of stay, with a greater accuracy of 97.17%.

Pritom et al. [15] conducted an extensive investigation into the forecasting of breast cancer recurrence, leveraging a suite of sophisticated classification algorithms, including SVM, Decision Trees, Naive Bayes, and the venerable C4.5 method. Through the use of efficient feature selection techniques, they hoped to increase the accuracy of each classifier. They employed the Info Gain characteristic with a ranker search engine as one such technique. The effectiveness of recurrence prediction was assessed by applying these algorithms on the Weka tool. Upon the unaltered dataset, void of any feature selection, the results unveiled a landscape where SVM distinguished itself with a remarkable precision of 75.75%. Meanwhile, the J48 secured an impressive 73.73% accuracy, while the Naïve Bayes classifier demonstrated notable proficiency, attaining a respectable accuracy of 67.17%. However, after properly implementing feature selection, SVM, C4.5, and Nave Bayes showed improvements of 1.52%, 2.52%, and 9.09%, respectively. This compelling evidence underscores feature selection's remarkable efficacy in elevating the classifiers' accuracy, validating its pivotal role in enhancing performance.

Dulhare et al. [7] constructed classification models to predict and categorize individuals with chronic kidney disease (CKD) using feature selection and the naive Bayes classifier. They used a feature selection technique known as the one R attribute selector to extract useful rules. The model's classification accuracy was evaluated using the best first search engine and the wrapper subset evaluator. Upon assimilating these methodologies into the Weka tool, the Naïve Bayes classifier achieved a notable accuracy threshold of 97.5%, accentuating the profound impact of their integration. This indicates how well the feature selection and classification strategy correctly identifies patients with and without CKD.

Artificial intelligence (AI) and deep learning have revolutionized various industries, including agriculture [16–23], education [24, 25], finance [26], healthcare [27–29] and other domains [30, 31]. In the field of healthcare, AI has shown tremendous promise in improving patient outcomes, enhancing diagnostics, and streamlining healthcare processes. With the ability to analyze vast amounts of data and identify complex patterns, AI-powered systems have opened new frontiers for early disease detection, personalized treatment plans [32], and overall healthcare efficiency. In healthcare, one of the areas where AI and deep learning have made significant advancements is in the early detection of diseases, including cancer [33]. Detecting cancer at an early stage is crucial for improving treatment success and patient survival rates. Kidney cancer, for example, often presents with few symptoms in its early stages, making early detection challenging. However, deep learning algorithms have proven to be effective in analyzing medical imaging data, such as CT scans and MRI images, to detect kidney tumors at their nascent stages [34].

III. Different Techniques Employed in the Realm of CKD

A. Artificial Neural Network (ANN)

The functioning of natural neurons served as the inspiration for artificial neural network (ANN), often known as a "neural network," which is widely used in practical applications. With changeable weights assigned to each link, it consists of interconnected nodes of artificial neurons, allowing for changes in their spatial layout during information transmission [4]. Being a learning algorithm, ANN can change its structure as it learns by taking in information from its internal and external environments [6]. The network has several layers for message propagation, including an input layer, hidden layers, and an output layer. One or more levels with different numbers of nodes may be included in the hidden layers. These layers are connected, and each node is tied to a certain weight. With ANN, input data is sent to the network to generate predictions while the network learns under supervision. A perceptron, which is the main working component of ANN, may divide datasets into two types. A perceptron is made up of a single node with corresponding weights. Its three key components are the connections between nodes, an adder for adding inputs, and activation functions that control the output.

B. K-Nearest Neighbor (KNN)

KNN is a nonparametric supervised learning technique that works well with both linear and nonlinear data [1]. It features a rapid training procedure and works especially well for big datasets. KNN determines the k objects that are closest to the place of interest or by majority voting. The selection of items hinges upon identifying the nearest class object, as determined by the minimal distance between the querying instance and the corresponding training example. According to Boukenze, et al. [4], KNN is the algorithm with the quickest model-building execution time. KNN considers k instances—x1, x2…—to forecast the class of a query Xn. Different distance metrics, such as the Euclidean, Manhattan, Minkowski, and Hamming distances, are used to determine which class is closest. This is how the distance formula is put together.

$$\text{Euclidean} = \sqrt{\sum_{i=1}^{n}(Xi - Yi)2}$$

$$\text{Manhattan} = \sum_{i=1}^{n} |Xi - Yi|$$

$$\text{Minkowski} = \sum_{k} (|Xi - Yi|)^{q} |)^{\frac{1}{q}}$$

C. Decision Tree (J48)

The Weka platform's C4.5 decision tree technique is implemented in Java as a J48 decision tree. It is a development of Ross Quinlan's original ID3 algorithm.

The top-down greedy search method is used by the J48 classification algorithm to build the tree. The final decision tree produced by J48 is made up of sorted branches, where the internal characteristics reflect potential outcomes based on the branching features and the leaf nodes represent the final class results. The separation between information gain and the splitting qualities is the foundation for the tree's creation. Data disorder and uncertainty are measured using entropy. The formula below can be used to determine the entropy of a random variable given a probability "p" and sample "S".

$$\text{Entropy(S)} = \sum_{i=1}^{n} (-pi\ log2(p_i))$$

The most informative aspect for choosing the best node in a decision tree is measured according to information gain. It measures the amount of entropy or uncertainty that is reduced due to dividing the data based on a particular attribute. We take into account the values of the attribute as "v" and the subset of sample "S" that corresponds to each value to determine the information gain for a given attribute "A" about a sample "S". Following are the steps for computing the information gain.

$$\text{Gain (S, A)} = \text{Entropy(S)} - \sum_{v \in \boldsymbol{Value}}^{\mathbf{n}} (\text{Entropy(Sv)} \frac{|\text{Sv}|}{|\text{S}|}$$

D. SVM: Support Vector Machine

One of the quintessential techniques in machine learning, renowned for its prowess in supervised learning, is the Support Vector Machine (SVM) [6]. It is famous for its high-performance abilities in classification, regression, analysis, and prediction tasks on datasets. In the feature space of the training data, SVM creates a distinct hyperplane that divides and categorizes the data points according to their positions about the hyperplane. In data mining, it is frequently used for problems involving regression and classification [1]. When compared to other classification algorithms, SVM, a widely used supervised learning technique, can produce accurate results. SVM strategically endeavors to minimize classification error by optimizing the margin between instances belonging to distinct classes, thereby culminating in an exceptional power. SVM uses the "kernel trick" to determine the separation between a data point and the hyperplane in a modified feature space without explicitly modifying the original features, which is one of its benefits.

E. Simple Bayes

A classification technique that applies the Bayes theorem under the feature independence presumption is known as the Naive Bayes classifier. It is a probabilistic technique that is frequently used in supervised classification issues in practical settings [1]. Naive Bayes has quick learning skills and works well in applications like diagnosis and prediction. The Nave Bayes algorithm works well with less training data, making it appropriate for classification jobs where data availability is constrained [35, 36]. Each record is predicted and associated with a particular class using the Naive Bayes classification algorithm, which calculates the likelihood that the record belongs to the target class. The most likely class is the one with the highest likelihood.

$$P(P/X) = \frac{P(P/X).P(Y)}{P(X)}$$

IV. Approach and Empirical Findings

The provided study revolves around two methodologies. The initial approach involves crafting a predictive model through the implementation of several feature curation tactics. The subsequent technique involves establishing a predictive model by harnessing an amalgamation of diverse classifiers.

A. Techniques for Choosing Features

A comprehensive classification model may be created by removing superfluous features from the data set and reducing the dimensionality of the feature. In this study investigators used the ranker search strategy with the Info Gain Attribute Evaluator feature selection technique to find the most relevant features. The Info Attribute Evaluator operates on a distinctive principle, evaluating an attribute's significance by comparing its information gain with respect to the classes. Moreover, it exhibits remarkable capability in appraising binary numeric attributes, eliminating the need for conventional feature discretization procedures. Moreover, the absent data could be distributed among alternative values based on their averages and commonly appearing values for a definite trait or a numeric feature. Alternatively, it might be considered as a distinct entity. The Attribute Evaluator for Information Gain showcases impressive adaptability, capable of identifying absent data and a range of attribute categories like nominal, temporal, numerical, unary, binary, and vacant minor attributes.

The ranker search technique can be exploited to obtain the rankings of attributes also evaluates each attribute by its specific evaluator in addition to attribute evaluators like information gain and Gini index. It possesses the capability to produce feature prioritization.

In this investigation, a supplementary layer of feature selection was implemented through the synergistic incorporation of a Wrapper Subset Evaluator, seamlessly melded with the precision-driven best first search methodology. This evaluative framework operates by leveraging sophisticated learning pattern to gauge the efficacy of attributes sets, fortified by a meticulous cross-validation process to affirm the soundness of the acquired insights. The detected attributes compass a wide array of characteristics, which consist of text, null nominal, undefined entries, time-related, relational, numeric, individual, paired, and categorical features. Furthermore, its discerning capabilities extend to the identification of attributes within Nominal, Binary, Date and Numeric classes, as well as the nuanced recognition of values within the Missing class domain.

In order to search the space of attribute subsets, first-best search makes use of a greedy hill climbing capacity supplemented with a backtracking capability. The maximum allowed number of consecutive non-improving nodes determines how much backtracking is allowed. The Best first search methodology offers a gamut of strategic pathways for exploration. It may embark upon its journey with an initial attributes set that is void, then propel forward, or alternatively, commence with a comprehensive set and elegantly traverse backward. Furthermore, it possesses the flexibility to initiate from any vantage point, seamlessly navigating in both forward and reverse directions. This dynamic approach encompasses the meticulous consideration of all possible single attribute

augmentation and eliminations at specific junctures and these modalities can be ingeniously combined to forge a tailored exploration strategy.

B. Ensemble Classifiers

The ensemble with the most heterogeneous classifier typically has the highest accuracy rate. Ensemble learning, a technique in the field of machine learning, revolves around creating multiple prediction models and blending their outputs to enhance the overall performance metrics of each algorithm. Employing an ensemble classifier is the optimal approach for rectifying errors that may arise from the underlying primary classifier [13]. In machine learning combination classification, it is currently increasingly usual to use many classifiers rather than just one. The advantage is that we can use two or more powerful categorization algorithms rather than just one. Hence, the resultant model will reach an elevated echelon of effectiveness and sophistication, equipping it to expertly discern and classify samples gathered from the training, cross-fold validation, and thoroughly examined testing datasets.

The ensemble classification model combines a number of classifiers, each of which affects the outcome in a unique way. These methodologies have ushered in a transformative evolution within the training process, molding classifier models that yield diverse categorization outcomes with unparalleled precision [12]. Ensemble methods' key advantage is that they combine individual classifier rules to provide predictions that are more accurate than those made by those rules alone. To improve prediction accuracy, the ensemble model approach mixes various individual classifiers.

The architecture of the proposed ensemble model is shown in Fig. 1.The proposed system's general methodology uses data from the UCI machine learning repository. For nominal and numeric variables within the dataset, any instances of missing values are meticulously imputed through the modal and mean values derived from the training data. Concurrent with this, the intricacy of features is finely tuned through the skilled utilization of feature selection methods. After feature reduction, we have the optimum feature subset, and the dataset was trimmed down for the proposed research. The reduced subset dataset consists of a few pertinent features. Following a rigorous process of individual attribute value feature selection, the foundational classifier was enriched with a cadre of prominent base classifiers, comprising K-Nearest Neighbors (KNN), J48 decision trees, Artificial Neural Networks (ANN), Naïve Bayes (NB) and Support Vector Machines (SVM). The data has been split into training and test sets. A training set of data was used to train the base classifier, a testing framework for evaluating and predicting illnesses. To improve the results, create a final forecast after combining the classifier using the ensemble voting method. The models' effectiveness is evaluated. Before this, the results were evaluated using ROC, F-Measure, recall, accuracy, and precision. The broad suggested system design is demonstrated in Fig. 1.

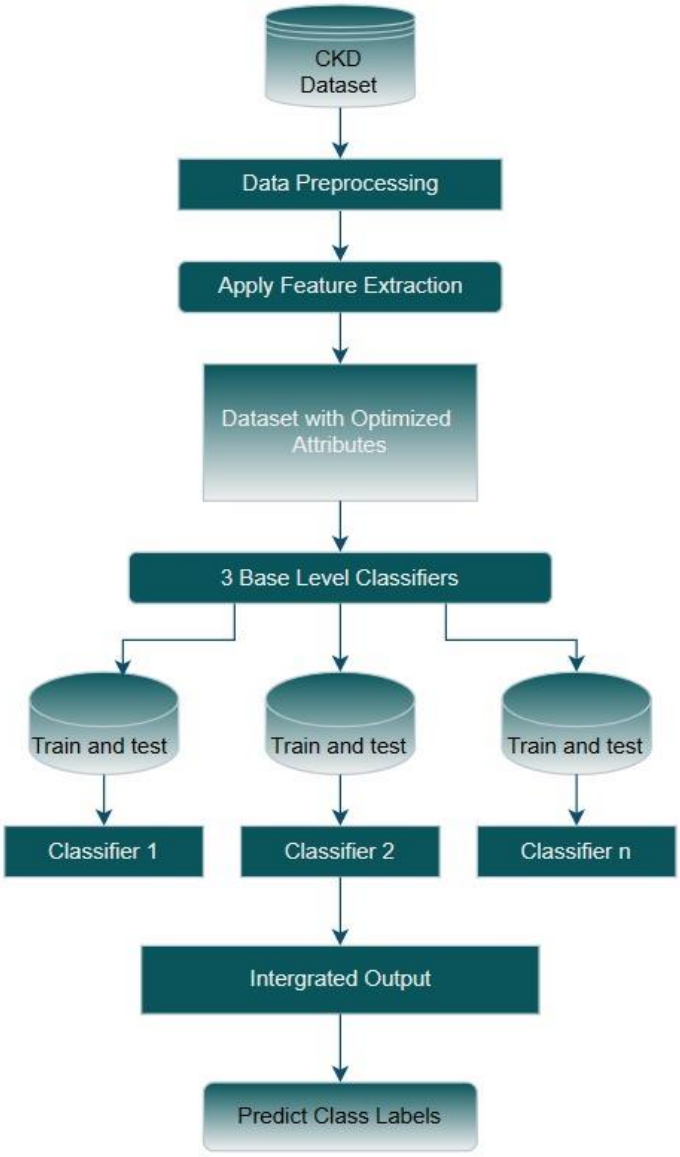

Fig. 1. The architecture of proposed ensemble model.

C. Dataset Exploited

The dataset from the UCI machine learning repository was acquired [11]. The collection contains 400 examples with 24 attribute values and 1 class attribute. These traits are included in the Table I below. The dataset contains 400 samples, 250 of which have CKD and 150 have not. It has the following characteristics: (14 numeric, 11 nominal) 24 + class = 25.

D. Various Performance Indicators Explored

By computing several performance indicators, a confusion matrix is used to gauge a classification algorithm's accuracy. It displays the classification model's correct and incorrect predictions with the dataset's actual values or intended results.

In this study, researchers took into account the predicted classes "CKD" and "not CKD." When predicting if a person has an illness, such as chronic kidney disease (CKD), "CKD" denotes that they have, but "not CKD" denotes that they do not.

TABLE I. SHOWS THE CHARACTERISTICS OF CHRONIC KIDNEY DISEASE [11]

Fields	Description
age	Age
bp	Blood-pressure
sg	Specific-gravity
al	Albumin
su	Sugar
rbc	Red blood cells
pc	Pus-cell
pcc	Pus-cell clumps
ba	Bacteria
bgr	Blood-glucose random
bu	Blood-urea
sc	Serum-creatinine
sod	Sodium
pot	Potassium
home	Hemoglobin
pcv	Packed-cell volume
wc	White blood-cell count
rc	Red blood-cell count
htn	Hypertension
dm	Diabetes-mellitus
cad	Coronary artery-disease
appet	Appetite
pe	Pedal-edema
ane	Anemia
class	Class

Also, investigators in this study employed the following indicators to assess how well our experiment performed using the confusion matrix in Table II.

True Positives (TP): These are cases that were accurately identified as positive, i.e., they had CKD when it was expected that they would.

- True Negative (TN): Instances that were appropriately identified as negative and were both predicted not to have CKD and really do not.

False Positive (FP) cases are those that were incorrectly identified as positive; they were expected to have CKD but didn't actually have it.

False Negatives (FN) are situations that were incorrectly categorized as negative; they were thought to be free of chronic renal disease but actually were.

- Accuracy: This pertains to the inherent ability of a classification algorithm to accurately anticipate and discern the underlying classes inherent within a given dataset. It is a metric that shows how accurately the classifiers assign the examples to the appropriate classes based on their projected class labels.

Accuracy is equal to (TP+TN)/(TP+TN+FP+FN).

TABLE II. CONFUSION MATRIX TABLE

Confusion Matrix				
	Positive	Negative	Target Value	
Positive	TP	FN	Positive Predictive value	$\frac{TP}{TP+FP}$
Negative	FP	TN	Negative Predictive value	$\frac{TN}{FN+TN}$
	$\frac{TP}{TP+FP}$ Precision	$\frac{TP}{TP+FN}$ Recall	Accuracy= TP+TN/TP+FP+FN+TN	

- Recall: It also goes by the name "sensitivity," and it gauges how well a classification algorithm can find pertinent examples.

Recall equals TP/ (TP+FN)

- Precision is a metric used by a classification system to evaluate how relevant the information that has been gathered is. It focuses on the percentage of retrieved instances that are relevant.

Precision is equal to TP / (TP+FP).

- F-Measure, sometimes referred to as the F-score, combines a test's recall and precision to determine how accurate it is. In order to assess the overall effectiveness of a classification system, it offers a balanced average of recall and precision.

F-Measure=2* (Recall + Precision) / (Recall*Precision)

- Receiver Operating Characteristics (ROC) Analysis: The Receiver Operating Characteristics (ROC) curve stands as a visual instrument of paramount significance for the appraisal of classification test efficacy. Presented graphically, it portrays the intricate interplay amidst the trade-offs between accurate identification and false positive occurrences across threshold variations. The ROC curve astutely encapsulates the nuanced balance between true positive and false positive outcomes, thereby facilitating a comprehensive comparison of distinct classification models. A pivotal gauge of model precision lies in the area under the ROC curve (AUC), a scalar metric ranging between 0 and 1. A value closer to 1 signifies a heightened model performance. Manifesting the capability to distinguish positive instances superiorly than negative instances, the AUC quantifies this discernment. On the ROC curve, the horizontal axis charts the false positive rate, while the vertical axis, often denoted as recall, delineates the true positive rate. A classification algorithm's performance and discriminatory ability are valuable revealed by the ROC curve.

TABLE III. DISPLAYS THE RESULTS OF CLASSIFIERS BOTH WITH AND WITHOUT FEATURE SELECTION

Categorization using feature selection techniques with and without	Precision	Recall	F-Measure	Accuracy
KNN absence of feature selection	0.965	0.966	0.965	96.60
KNN with JofeGainAttributeEval and ranker (chosen 20 Attributes)	0.977	0.971	0.978	97.74
KNN and Ranker with 15 Attributes)	0.97	0.97	0.97	97
Best first search engine with WrannerSubsetExal (Picked 8 attributes)	0.98	0.98	0.98	98
J48 absence of feature selection	0.956	0.956	0.957	95.76
JofeGainAttributeExal using a ranker (20 Attributes chosen)	0.977	0.978	0.977	97.74
J48 with JofeGajoAttributeEval and ranker (15 Attribute chosen)	0.977	0.978	0.977	97.76
Best first search engine 48 WrapperSubsetExal (7 attributes chosen)	0.973	0.979	0.978	97.77
ANN absence of feature selection	0.968	0.968	0.968	96.65
ANN and ranker with JofeGainAttributeEval (20 Attributes chosen)	0.971	0.97	0.97	97
ANN and ranker with JofeGainAttributeExal (15 Attribute chosen)	0.976	0.965	0.975	97.55
Best first search engine with WrannerSubsetExal and ANN (8 attributes chosen)	0.975	0.979	0.976	97.78
NB absence of feature selection	0.941	0.935	0.936	93.25
NB and ranker with Attributes chosen)	0.942	0.938	0.938	93.45
NB and ranker with JofeGainAstributeEval (15 Attributes chosen)	0.956	0.953	0.952	95.12
Best first search engine with WrannerSubsetEval and NB (9 attributes chosen)	0.971	0.969	0.970	97
SVM absence of feature selection	0.968	0.969	0.968	96.76
SVM and ranker with InfoGainAttributeEval 20 Attribute chosen)	0.973	0.938	0.973	97.17
SVM and ranker with InfoGainAttributeEval, 15 Attributes chosen)	0.978	0.979	0.978	97.69
Best first search engine SVM with WrapperSubsetEval and SVM (8 attributes chosen)	0.971	0.972	0.973	97.12

In pursuit of enhanced precision in predicting Chronic Kidney Disease (CKD), a meticulous endeavor is made to curtail dataset dimensionality through strategic feature selection approaches. The Info Gain attribute evaluator synergistically combines forces with the ranker search engine, while the WrapperSubsetEval harnesses the prowess of the Best first search engine. By juxtaposing the outcomes against the original dataset, these methodologies ingeniously sculpt a novel dataset boasting reduced dimensions. The culmination of classifier results, with and without the application of feature selection techniques, finds concise summation in the depicted Table III above.

The dataset dimension for each classifier was 20 attributes as a result of the initial feature selection approach, which combined the infoGainAttributeEval evaluator with the ranker search engine. The infoGainAttributeEval evaluator and ranker search engine was then used in a second feature reduction procedure to further reduce the dimensionality, yielding a dataset characterized by 15 attributes per classifier, with the integration of the third feature selection strategy. The Best First search engine with the WrapperSubsetEval evaluator, resulting in a dataset dimension of eight attributes.

Without feature selection, all 25 characteristics were used in the K-nearest neighbour (KNN) classifier. The dataset dimensions for J48, ANN, Naive Bayes, and SVM classifiers were reduced to seven attributes for J48, ANN, and Naive Bayes, and eight attributes for SVM, using the WrapperSubsetEval evaluator with the Best First search engine.

Fig. 2 exhibits the empirical revelations pertaining to each classifier, juxtaposing their performance both in the presence and absence of the feature selection methodology. A comprehensive depiction of the classifiers' performance metrics, encompassing precision, recall, F-Measure, and accuracy, is meticulously presented within the same figure. In order to assess the CKD and non-CKD cohorts, discerning weighted averages were invoked as a robust analytical framework.

According to Fig. 2 findings, the ANN classifier's CKD prediction accuracy was highest when feature selection was combined with the WrapperSubsetEval and Best First search engines. The top-ranked average figures for precision, recall, F-Measure, and obtained using this method, which chose eight attributes from the original 25. The ANN classifier achieved a prediction accuracy of 97.78% for CKD without feature selection. The accuracy of the KNN classifier dropped to 97.55% when utilizing the InfoGainAttributeEval with ranker feature selection approach, demonstrating that the removal of some attributes had a detrimental effect on accuracy. As opposed to the accuracy of the normal dataset, which was 96.65%, the accuracy increased to 97% with the selection of 20 attributes.

When employing feature selection with InfoGainAttributeEval and ranker, the J48 classifier had the highest respectable accuracy in CKD prediction, as seen by the values in Fig. 3. The J48 classifier attained an accuracy of 97.75% approximately for both 15 and 20 specified attributes. The J48 classifier, on the other hand, showed a lower accuracy of 95.76% when predicting CKD without feature selection. The dataset was reduced to seven attributes using feature selection, WrapperSubsetEval, and the Best First search engine, and the J48 classifier showed a 97.77% accuracy rate compared to using the entire dataset with all 25 attributes; this accuracy rate was more significant.

The KNN classifier's accuracy rates for predicting CKD differed depending on the feature selection techniques used.

The KNN classifier acquired an accuracy rate of 96.60% in CKD prediction without requiring feature selection. The accuracy increased to 97.74% after implementing the InfoGainAttributeEval with a ranker feature selection approach and choosing 20 characteristics. However, when choosing 15 attributes utilizing the InfoGainAttributeEval with a ranker feature selection approach, the accuracy rate dropped to 97%. When dimensionality was reduced using the WrapperSubsetEval with the Best initial search engine feature selection method, the ANN classifier's accuracy rate on the CKD dataset increased. With eight qualities chosen, the accuracy rate was 98%.

The Naive Bayes classifier's accuracy rates for predicting CKD differed depending on the feature selection techniques used. The Nave Bayes classifier attained an accuracy rate of 93.25% in CKD prediction without requiring feature selection. The accuracy increased slightly to 93.45% after applying the InfoGainAttributeEval with a ranker feature selection approach and choosing 20 attributes. The accuracy rate increased to 95.12% while utilizing the InfoGainAttributeEval with a ranker feature selection approach and using 15 points. On the CKD dataset, dimensionality reduction utilizing the WrapperSubsetEval with the Best initial search engine feature selection strategy led to the Naive Bayes classifier's most excellent accuracy rate, nine fields were chosen, and the accuracy rate was 98%.

Regarding predicting CKD, the SVM classifier's accuracy rates they differed according to the feature selection techniques used. Without feature selection, the SVM classifier only managed to predict CKD with an accuracy rate of 96.76%.

Graphical demonstration with and without feature selection, the precision, recall, and F-measures of chronic renal illness are compared in Fig. 2.

The accuracy increased to the highest rate of 97.17% after implementing the InfoGainAttributeEval with a ranker feature selection approach and choosing 20 characteristics. The accuracy rate amplified to 97.69% when 15 attributes were selected using the InfoGainAttributeEval with ranker feature selection approach. On the CKD dataset, dimensionality reduction utilizing the WrapperSubsetEval with the Best initial search engine feature selection strategy led to the SVM classifier's most excellent accuracy rate. 8 attributes were chosen, and the accuracy percentage was 97%. This rate, meanwhile, needed to be more accurate with the rate attained with the 20-dimensional dataset.

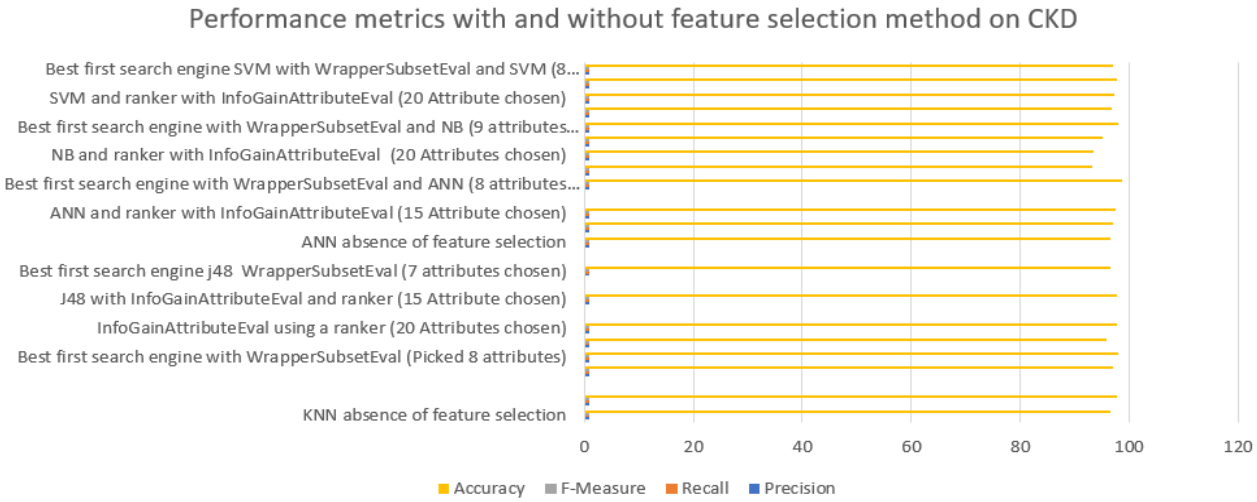

Fig. 2. Comparison of feature selection, precision, recall and F-measure of chronic renal illness.

TABLE IV. EVALUATION OF CLASSIFIERS USING ENSEMBLE METHODS AND FEATURE SELECTION

Classifier	Precision	Recall	F-Measure	Accuracy
KNN	0.965	0.966	0.965	96.60
KNN after FS	0.98	0.98	0.980	98
J48	0.956	0.956	0.957	95.76
J4N after FS	0.977	0.979	0.978	97.77
ANN	0.958	0.968	0.968	96.55
ANN after FS	0.977	0.979	0.976	97.78
NR	0.971	0.969	0.936	93.25
NB after FS	0.968	0.969	0.970	97
SVM	0.968	0.969	0.968	96.79
SVM after FS	0.971	0.972	0.973	97.12
Essemble model	0.985	0.986	0.985	98.85

While evaluating the ensemble classifier model on the CKD dataset, an exhaustive scrutiny of performance metrics was conducted, encompassing accuracy, precision, recall, F-Measure, actual positive rate, and ROC comparisons. These comprehensive assessments were meticulously juxtaposed with the individual classifiers' corresponding outcomes. Impressively, the ensemble model showed a high accuracy rate in CKD prediction.

Performances of classifiers with and without methods for feature selection and assembly are shown in Table IV.

The envisaged ensemble framework outclassed other standalone base classifiers and was combined with most of our suggested feature selection techniques. Various performance metrics were used to compare the ensemble model to heterogeneous base classifiers with and without feature selection. The ensemble model used lower dimensions acquired by feature selection techniques, which helped reduce training time and computing expenses. By cutting costs and execution time, feature selection increased accuracy. Fig. 3 shows that the ensemble model demonstrated the highest accuracy level for the CKD dataset compared to the individual base classifiers. It obtained 98.85% accuracy, 0.985% precision, 0.986% recall, and 0.985% F-Measure rates.

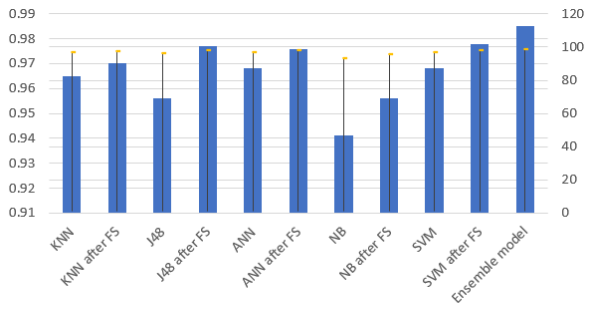

Fig. 3. Demonstrates the accuracy assessment of different classifiers.

Fig. 4 presents an insightful comparative analysis, meticulously evaluating the base classifier's level of correctness, exactness, and ability to retrieve relevant instances, both in the presence and absence of feature selection. Discernible in Fig. 4, the ensemble model decidedly outperformed the other classifiers. A comprehensive comparison of the ensemble's performance vis-à-vis that of the remaining classifiers for the dataset about chronic renal illness showcased the ensemble's profound superiority.

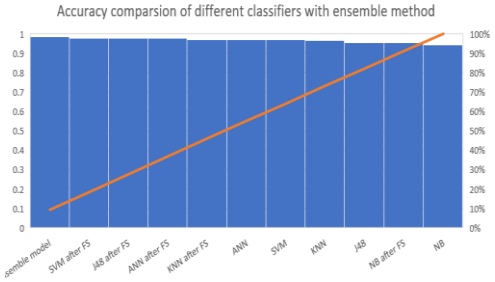

Fig. 4. Performance comparison of all the classifiers exploited in this study using a Pareto chart plot.

V. Conclusion

In the pursuit of bolstering classifier precision, the present study diligently harnessed an arsenal of feature selection methodologies and strategically integrated ensemble models within the domain of the CKD dataset. Each classifier was guided by a distinct feature selection evaluator, encompassing the sophisticated InfoGainAttributeEval, the meticulous Ranker search platform, and the astute WrapperSubsetEval feature seamlessly integrated with the proficient Adept Best-First Search mechanism. The strategic application of these techniques was intricately woven into both the ensemble model and the proposed feature selection approach, amplifying the precision of the machine learning classifiers. The performance of K-Nearest Neighbors (KNN), J48, Artificial Neural Networks (ANN), Naïve Bayes (NB), and Support Vector Machine (SVM) classifiers was meticulously juxtaposed across the CKD dataset, and a refined subset crafted through WrapperSubsetEval, featuring the Best First Search Engine and InfoGainAttributeEval as feature selection evaluators. The experimental findings showed that increasing the dataset's dimension decreased the classifiers' accuracy. In particular, the accuracy of ANN classification using WrapperSubsetEval and the Best First search engine on the condensed dataset was 97.78%, exceeding the accuracy attained using the primary dataset and various approaches to feature selection. Contrasted against the unaltered dataset and alternative feature selection methodologies, the condensed dataset derived through the astute combination of InfoGainAttributeEval and the ranker search engine exhibits a remarkable accuracy pinnacle of 97.77% in the realm of J48 classification. With WrapperSubsetEval and the Best initial search engine, the KNN classification accuracy on the condensed dataset reached 98%, the most fantastic accuracy of any approach. With the help of WrapperSubsetEval and the Best first search engine, Naive Bayes classification was 97% accurate on the smaller dataset, outperforming the original dataset and other feature selection techniques. In parallel to the accuracy achieved using the initial dataset and various feature selection techniques, the SVM classifier achieved an impressive precision of 97.12% on the improved dataset generated by employing InfoGainAttributeEval in combination with the ranker search engine. These techniques engendered a discernible reduction in the false positive rate, concomitantly fostering an augmentation in the actual positive rate. Furthermore, an encompassing enhancement encompassed performance metrics alternatively, terms like accuracy, sensitivity, harmonic mean of precision and recall, and exact positive ratio. The ensemble techniques proposed demonstrated superior performance in classifying and predicting CKD on the provided dataset, as evidenced by the experimental results of these ensemble methods. On the condensed CKD dataset, the ensemble classification's accuracy exceeded that of the individual base classifiers by 98.85%. Looking to the future, the ensemble techniques we proposed showcased superior performance in classifying and predicting CKD on the provided dataset, as demonstrated by the impressive accuracy rate of 98.85%. In the upcoming research, further exploration of ensemble methods and feature selection techniques can be pursued to enhance CKD prediction models. Additionally, investigating the

generalizability of these techniques to other medical diagnoses and datasets would be a promising direction for future research in the domain of machine learning and healthcare.

ACKNOWLEDGMENT

King Faisal University: Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia, under the Project GRANT4,487.

REFERENCES

[1] Chandel, K.; Kunwar, V.; Sabitha, S.; Choudhury, T.; Mukherjee, S. A Comparative Study on Thyroid Disease Detection Using K-Nearest Neighbor and Naive Bayes Classification Techniques. CSI Transactions on ICT 2016, 4, 313–319, doi:10.1007/S40012-016-0100-5.

[2] Padmanaban, K.; Journal, G.P.-I.; 2016, undefined Applying Machine Learning Techniques for Predicting the Risk of Chronic Kidney Disease. sciresol.s3.us-east-2.amazonaws KRA Padmanaban, G ParthibanIndian Journal of Science and Technology, 2016•sciresol.s3.us-east-2.amazonaws … 9, doi:10.17485/ijst/2016/v9i29/93880.

[3] Majid, M.; Gulzar, Y.; Ayoub, S.; Khan, F.; Reegu, F.A.; Mir, M.S.; Jaziri, W.; Soomro, A.B. Enhanced Transfer Learning Strategies for Effective Kidney Tumor Classification with CT Imaging. International Journal of Advanced Computer Science and Applications 2023, 14, 2023, doi:10.14569/IJACSA.2023.0140847.

[4] Boukenze, B.; Haqiq, A.; Mousannif, H. Predicting Chronic Kidney Failure Disease Using Data Mining Techniques. Lecture Notes in Electrical Engineering 2017, 397, 701–712, doi:10.1007/978-981-10-1627-1_55.

[5] Dhiman, P.; Bonkra, A.; Kaur, A.; Gulzar, Y.; Hamid, Y.; Mir, M.S.; Soomro, A.B.; Elwasila, O. Healthcare Trust Evolution with Explainable Artificial Intelligence: Bibliometric Analysis. Information 2023, Vol. 14, Page 541 2023, 14, 541, doi:10.3390/INFO14100541.

[6] Polat, H.; Danaei Mehr, H.; Cetin, A. Diagnosis of Chronic Kidney Disease Based on Support Vector Machine by Feature Selection Methods. J Med Syst 2017, 41, doi:10.1007/S10916-017-0703-X.

[7] Dulhare, U.; Conference, M.A.-2016 I.I.; 2016, undefined Extraction of Action Rules for Chronic Kidney Disease Using Naïve Bayes Classifier. ieeexplore.ieee.orgUN Dulhare, M Ayesha2016 IEEE International Conference on Computational Intelligence, 2016•ieeexplore.ieee.org.

[8] Radhakrishnan, J.; Mohan, S. KI Reports and World Kidney Day. Kidney Int Rep 2017, 2.

[9] Sedighi, Z.; Ebrahimpour-Komleh, H.; Mousavirad, S.J. Featue Selection Effects on Kidney Desease Analysis. In Proceedings of the 2nd International Congress on Technology, Communication and Knowledge, ICTCK 2015; 2016.

[10] Khan, F.; Ayoub, S.; Gulzar, Y.; Majid, M.; Reegu, F.A.; Mir, M.S.; Soomro, A.B.; Elwasila, O. MRI-Based Effective Ensemble Frameworks for Predicting Human Brain Tumor. Journal of Imaging 2023, Vol. 9, Page 163 2023, 9, 163, doi:10.3390/JIMAGING9080163.

[11] National Kidney Foundation Global Facts: About Kidney Disease Available online: https://www.kidney.org/kidneydisease/global-facts-about-kidney-disease (accessed on 12 October 2023).

[12] Bashir, S.; Qamar, U.; Khan, F.H.; Javed, M.Y. MV5: A Clinical Decision Support Framework for Heart Disease Prediction Using Majority Vote Based Classifier Ensemble. Arab J Sci Eng 2014, 39, doi:10.1007/s13369-014-1315-0.

[13] Bashir, S.; Qamar, U.; Khan, F.H.; Naseem, L. HMV: A Medical Decision Support Framework Using Multi-Layer Classifiers for Disease Prediction. J Comput Sci 2016, 13, doi:10.1016/j.jocs.2016.01.001.

[14] Khajehali, N.; Alizadeh, S. Extract Critical Factors Affecting the Length of Hospital Stay of Pneumonia Patient by Data Mining (Case Study: An Iranian Hospital). Artif Intell Med 2017, 83, doi:10.1016/j.artmed.2017.06.010.

[15] Pritom, A.I.; Munshi, M.A.R.; Sabab, S.A.; Shihab, S. Predicting Breast Cancer Recurrence Using Effective Classification and Feature Selection Technique. In Proceedings of the 19th International Conference on Computer and Information Technology, ICCIT 2016; 2017.

[16] Gulzar, Y. Fruit Image Classification Model Based on MobileNetV2 with Deep Transfer Learning Technique. Sustainability 2023, 15, 1906.

[17] Dhiman, P.; Kaur, A.; Balasaraswathi, V.R.; Gulzar, Y.; Alwan, A.A.; Hamid, Y. Image Acquisition, Preprocessing and Classification of Citrus Fruit Diseases: A Systematic Literature Review. Sustainability 2023, Vol. 15, Page 9643 2023, 15, 9643, doi:10.3390/SU15129643.

[18] Mamat, N.; Othman, M.F.; Abdulghafor, R.; Alwan, A.A.; Gulzar, Y. Enhancing Image Annotation Technique of Fruit Classification Using a Deep Learning Approach. Sustainability 2023, 15, 901.

[19] Gulzar, Y.; Ünal, Z.; Akta¸s, H.A.; Mir, M.S. Harnessing the Power of Transfer Learning in Sunflower Disease Detection: A Comparative Study. Agriculture 2023, Vol. 13, Page 1479 2023, 13, 1479, doi:10.3390/AGRICULTURE13081479.

[20] Aggarwal, S.; Gupta, S.; Gupta, D.; Gulzar, Y.; Juneja, S.; Alwan, A.A.; Nauman, A. An Artificial Intelligence-Based Stacked Ensemble Approach for Prediction of Protein Subcellular Localization in Confocal Microscopy Images. Sustainability 2023, Vol. 15, Page 1695 2023, 15, 1695, doi:10.3390/SU15021695.

[21] Gulzar, Y.; Hamid, Y.; Soomro, A.B.; Alwan, A.A.; Journaux, L. A Convolution Neural Network-Based Seed Classification System. Symmetry (Basel) 2020, 12, 2018.

[22] Malik, I.; Ahmed, M.; Gulzar, Y.; Baba, S.H.; Mir, M.S.; Soomro, A.B.; Sultan, A.; Elwasila, O. Estimation of the Extent of the Vulnerability of Agriculture to Climate Change Using Analytical and Deep-Learning Methods: A Case Study in Jammu, Kashmir, and Ladakh. Sustainability 2023, Vol. 15, Page 11465 2023, 15, 11465, doi:10.3390/SU151411465.

[23] Albarrak, K.; Gulzar, Y.; Hamid, Y.; Mehmood, A.; Soomro, A.B. A Deep Learning-Based Model for Date Fruit Classification. Sustainability 2022, 14.

[24] Sahlan, F.; Hamidi, F.; Misrat, M.Z.; Adli, M.H.; Wani, S.; Gulzar, Y. Prediction of Mental Health Among University Students. International Journal on Perceptive and Cognitive Computing 2021, 7, 85–91.

[25] Hamid, Y.; Elyassami, S.; Gulzar, Y.; Balasaraswathi, V.R.; Habuza, T.; Wani, S. An Improvised CNN Model for Fake Image Detection. International Journal of Information Technology 2022 2022, 1–11, doi:10.1007/S41870-022-01130-5.

[26] Gulzar, Y.; Alwan, A.A.; Abdullah, R.M.; Abualkishik, A.Z.; Oumrani, M. OCA: Ordered Clustering-Based Algorithm for E-Commerce Recommendation System. Sustainability 2023, Vol. 15, Page 2947 2023, 15, 2947, doi:10.3390/SU15042947.

[27] Gulzar, Y.; Khan, S.A. Skin Lesion Segmentation Based on Vision Transformers and Convolutional Neural Networks—A Comparative Study. Applied Sciences 2022, Vol. 12, Page 5990 2022, 12, 5990, doi:10.3390/APP12125990.

[28] Khan, S.A.; Gulzar, Y.; Turaev, S.; Peng, Y.S. A Modified HSIFT Descriptor for Medical Image Classification of Anatomy Objects. Symmetry (Basel) 2021, 13, 1987.

[29] Alam, S.; Raja, P.; Gulzar, Y. Investigation of Machine Learning Methods for Early Prediction of Neurodevelopmental Disorders in Children. Wirel Commun Mob Comput 2022, 2022.

[30] Ayoub, S.; Gulzar, Y.; Reegu, F.A.; Turaev, S. Generating Image Captions Using Bahdanau Attention Mechanism and Transfer Learning. Symmetry (Basel) 2022, 14, 2681.

[31] Hanafi, M.F.F.M.; Nasir, M.S.F.M.; Wani, S.; Abdulghafor, R.A.A.; Gulzar, Y.; Hamid, Y. A Real Time Deep Learning Based Driver Monitoring System. International Journal on Perceptive and Cognitive Computing 2021, 7, 79–84.

[32] Hamid, Y.; Wani, S.; Soomro, A.B.; Alwan, A.A.; Gulzar, Y. Smart Seed Classification System Based on MobileNetV2 Architecture. In Proceedings of the 2022 2nd International Conference on Computing and Information Technology (ICCIT); IEEE, 2022; pp. 217–222.

[33] Mehmood, A.; Gulzar, Y.; Ilyas, Q.M.; Jabbari, A.; Ahmad, M.; Iqbal, S. SBXception: A Shallower and Broader Xception Architecture for Efficient Classification of Skin Lesions. Cancers 2023, Vol. 15, Page 3604 2023, 15, 3604, doi:10.3390/CANCERS15143604.

[34] Gulzar, Y.; Alkinani, A.; Alwan, A.A.; Mehmood, A. Abdomen Fat and Liver Segmentation of CT Scan Images for Determining Obesity and Fatty Liver Correlation. Applied Sciences 2022, Vol. 12, Page 10334 2022, 12, 10334, doi:10.3390/APP122010334.

[35] Anand, V.; Gupta, S.; Gupta, D.; Gulzar, Y.; Xin, Q.; Juneja, S.; Shah, A.; Shaikh, A. Weighted Average Ensemble Deep Learning Model for Stratification of Brain Tumor in MRI Images. Diagnostics 2023, Vol. 13, Page 1320 2023, 13, 1320, doi:10.3390/DIAGNOSTICS13071320.

[36] Ayoub, S.; Gulzar, Y.; Rustamov, J.; Jabbari, A.; Reegu, F.A.; Turaev, S. Adversarial Approaches to Tackle Imbalanced Data in Machine Learning. Sustainability 2023, Vol. 15, Page 7097 2023, 15, 7097, doi:10.3390/SU15097097.