International Journal of Intelligent Information Systems
2022; 11(4): 51-64
http://www.sciencepublishinggroup.com/j/ijiis
doi: 10.11648/j.ijiis.20221104.11
ISSN: 2328-7675 (Print); ISSN: 2328-7683 (Online)

Ensemble Learning Improvement of Clinical Diagnoses of Knee Osteoarthritis Risk in Adults

Olayemi Olufunke Catherine[1], Olasehinde Olayemi Oladimeji[2], Alowolodu Olufunso Dayo[3], Osho Patrick Olarewaju[4]

[1]Department of Computer Science, Joseph Ayo Babalola University, Ikeji-Arakeji, Nigeria

[2]Department of Computer Science, Federal Polytechnic, Ile Oluji, Nigeria

[3]Department of Cyber Security, Federal University of Technology, Akure, Nigeria

[4]Department of Hematology & Immunology, University of Medical Sciences, Ondo, Nigeria

Email address:

ocolayemi@jabu.edu.ng (O. O. Catherine), olaolasehinde@fedpolel.edu.ng (O. O. Oladimeji), odalowolodu@futa.edu.ng (A. O. Dayo), droshopo@unimed.edu.ng (O. P. Olarewaju)

To cite this article:

Olayemi Olufunke Catherine, Olasehinde Olayemi Oladimeji, Alowolodu Olufunso Dayo, Osho Patrick Olarewaju. Ensemble Learning Improvement of Clinical Diagnoses of Knee Osteoarthritis Risk in Adults. *International Journal of Intelligent Information Systems*. Vol. 11, No. 4, 2022, pp. 51-64. doi: 10.11648/j.ijiis.20221104.11

Received: May 28, 2022; **Accepted**: June 27, 2022; **Published**: July 26, 2022

Abstract: *Background:* Knee Osteoarthritis (KOA) is a deteriorating disease that affects human knee joints leading to impaired quality of life with no curative treatments. Timely detection of KOA will guarantee its good management, prevent cartilage impairment and reduce its rate of progression. To heighten its early detection. *Objective:* This study developed a machine learning ensemble model that improves early clinical diagnosis of the risk of KOA in Adults. *Method:* The diagnostic results of three machine learning diagnostic models were combined with two ensemble methods proposed to improve the diagnosis of KOA risks. KOA patient dataset used for the modeling of the diagnostic models was obtained from the Federal Medical Hospital located in Ido-Ekiti, Nigeria. *Results and Conclusion:* The diagnostic result of the base diagnoses models shows higher accuracy than similar recently reviewed research in the literature. Diagnoses results of the two ensemble models confirm their abilities to improve the results of the base models. From the comparison of the diagnoses of the ensemble methods, the Multi Response Linear Regression model leads with 97.77% followed by the Majority Voting model with 96.54% diagnostic accuracy. The Statistical tests employed in this study, validated the ranking of the results recorded by each of the diagnostic models.

Keywords: Osteoarthritis, Clinical-Diagnoses, Ensemble Learning, Computational Intelligence, Improve Diagnoses

1. Introduction

Knee Osteoarthritis (KOA), also referred to as a "wear and tear" disease, is a deteriorating sickness that usually affects the human knee joints. It occurs when the slippery cartilage in the bone joints responsible for frictionless movement wears completely and causes bones to touch one another. Based on the etiology of the disease, there are two types of Osteoarthritis (OA); primary OA (idiopathic or non-traumatic) and the secondary OA (due to trauma or mechanical disorder). OA is solely a deteriorating disease of the cartilage, although the latest evidence has shown that OA is a complex entity involving different risk factors such as trauma, mechanical forces, inflammation, biochemical reactions, and metabolic imbalances [1]. KOA is the most common form of arthritis and one of the foremost causes of disability globally, affecting 3.8% of the global population [2]. It causes painful joint locking, which usually impairs the affected individual's daily functional activities and this frequently happens among middle-aged and elderly [3]. KOA is more pronounced in women than in men [2]. It was estimated that more than 27 million Americans have this condition, which primarily affects people who are 40 years of age or older [4]. A study carried out in a Nigerian hospital shows a high prevalence of OA with a high incidence in women with the knee joint being mostly affected with low involvement in the hand joints [5].

Findings of Palazzo et al. show that the risk of knee KOA reduces by 50%, with every 5 kg/m^2 reduction in weight [6-7]. Zheng and Chen, observed that overweight and obesity were significantly associated with higher KOA risks [8]. The Centre for Disease Control has estimated that nearly 1 in every two people develop KOA symptoms by age 85 [9]. According to, Symptoms of KOA are often developed slowly and worsen over time and this invariably leads to osteophyte formation, weakening of the periarticular muscles, slackness of ligaments, and synovial effusion. Signs and symptoms of KOA include pains in the affected joints during or after movement, noticeable joint stiffness upon awakening or after being inactive, grating sensation, bone spurs, and swelling [10-11]. Due to poor blood flow and innervation, the cartilage production of pain or inflammation does not occur, at least in the initial stages of the disease. Therefore, pain is derived solely from changes to the joints' non-cartilaginous parts like the joint capsule, subchondral bone, ligaments, and periarticular muscles [12]. Esser and Bailey Infer pain relief as the primary treatment for KOA's patients [13]. Knee replacement surgery is the only effective cure for KOA at its advanced stage. Early traces of cartilage injury could enable KOA detection at the potentially reversible stage when bone damage has not occurred [14]. According to Tiulpin et al, early diagnosis of KOA will ensure its proper management, prolong healthy patient-years, prevent cartilage from falling apart to slow down its progression and reduce its effect on future disability [15]. Physical examination of the knee for possible signs of swelling and checking for the contact between the tibia and femur bones in the X-ray image are some of the methods of diagnosing KOA. X-ray images of the affected knee cannot detect KOA in its early stage. Plain radiography is not sensitive enough to detect initial KOA changes [15]. Whenever the x-ray image examination is inconclusive to diagnose KOA, Magnetic resonance imaging (MRI) could be taken to view the presence of soft tissues; cartilage, ligament, and meniscus around the knee bones. The X-ray cannot view soft tissues around bones but can indirectly represent soft tissue integrity by assessing the joint space width. If this disease is detected at an early stage, a lot of pain and agony could be prevented. Medical diagnoses of the disease could be out of the reach of low-level income earners. The possibility of its clinical diagnoses remained the only hope for low-income earners. Clinical diagnosis is the method of detection of a disease or condition based on the symptoms and signs exhibited by a patient. Pass medical history of several KOA patients can be used to model a diagnoses model to diagnose new cases of the disease. This has informed the proposed hybridized system comprising of machine learning and ensemble learning for the early clinical detection of the risk of KOA.

Machine learning (ML) algorithms have been used to analyze and generate intelligence to discover hidden and potentially useful patterns from medical datasets for the correct prescription and treatment of diagnosed diseases [16]. The report of Esteva et al. put at par the ability of ML intelligence to diagnose and classify skin cancer at the same level of competence as a dermatologist [17]. The predictions and diagnoses of the ML models could be improved with the use of Ensemble Learning. Ensemble learning is a mathematical and statistical procedure that combines the predictions of a set of base learner models to give superior and improved predictive accuracy. Figure 1 shows the ensemble learning method, which comprises the predictions of the evaluation of dataset X on model 1, model 2…model n, also known as the base models. These predictions are combined together using ensemble learners (combiners) to produce the final prediction/diagnosis Y.

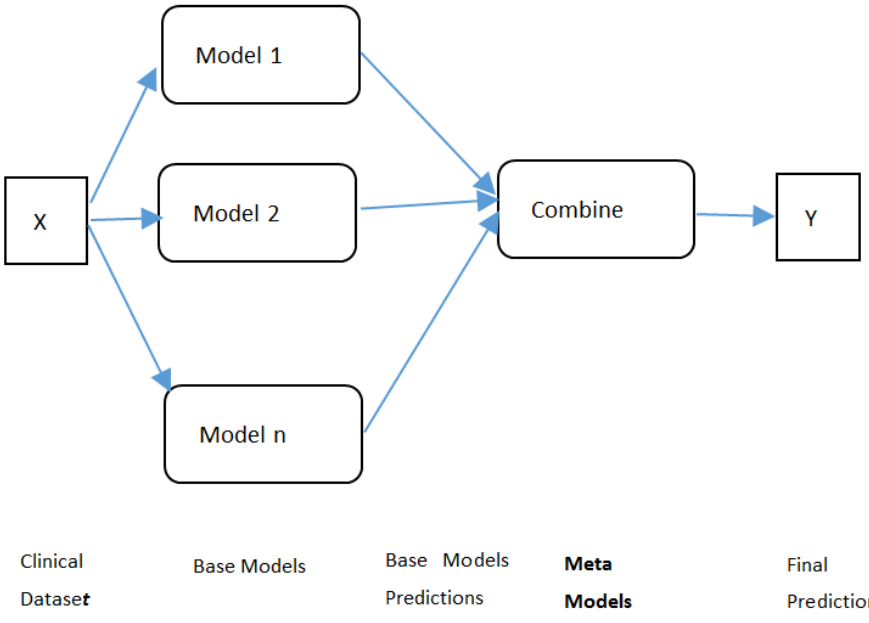

Figure 1. *Ensemble Learning Method.*

This research applies ensemble learning for optimal clinical diagnostic accuracy improvement of KOA risk in Adults, five

diagnostic models, comprising of Logistic Regression (LR), Support Vector Machine (SVM), C4.5 Decision Tree (DT), Naïve Bayes (NB), and K-Nearest Neighbors (KNN) were selected for our base models. The majority voting and stacking with the Multi Response Linear Regression (MLR) were selected for the ensemble meta-models to combine the five base models' diagnoses to build the ensemble models. Kendall's tau_b Correlations ranking and Paired differences for Ranking Statistical tests were used to validate the diagnostic results of the base and the ensemble models. A dataset of patients' information from the Federal Medical Centre, Ido-Ekiti, Nigeria, that contains the KOA risk factors, was used for this research. The results affirm the knack of ensemble learning to optimize the diagnostic accuracy of the risk of KOA. Improved accuracy in the diagnosis of KOA will result in early and proper management and handling of patients with KOA. This study has been able to:

i. Presents a novel ensemble learning diagnostic intelligence model for the clinical diagnoses of the risk of KOA. The results confirm the knack of ensemble learning techniques to improve the accuracy of the computational intelligence of the clinical diagnoses of KOA.
ii. The statistical tests carried out validated the diagnostic results of the computational intelligence models.

The remaining of this report is as follows; the literature reviews on ML's techniques for diagnosing KOA and applying ensemble learning to the medical diagnosis improvements was presented in section two. Section three described the dataset pre-processing method, the modeling techniques, and the performance evaluation metrics employed for this study. Experimental results and discussion is presented in section 4, and section 5 provides the conclusion of this research.

2. Review of Related Literature

The KOA disease had been known over time as a disorder that had impaired many lives with research showing that early detection could avert so many disabilities. Many pieces of research had been conducted by both academia and the health sector in finding solutions to its prevention, early detection, and cure. Machine learning and Ensemble learning had been used for the predictions and diagnosis of several diseases. The works of Olasehinde et al. ascertained the efficiency of the stacked ensemble to harness several ML diagnosis models' strengths to improve their diagnostic accuracy [18]. Olasehinde and Olayemi applied a stacked ensemble to improve the diagnosis of Lower Respiratory Tract infection (LRTI) in pediatric patients [16]. The report of Verma and Pal presented improved diagnosis and prediction of basic treatment for skin diseases [19]. In Rajaraman et al. a stacked ensemble was applied to improve Tuberculosis detection in chest radiographs [20]. Khalid applied ensemble learning to improve the diagnostic accuracy of heart disease [21]. Yoo et al. presented research work on the prediction of rheumatoid arthritis, using a machine learning algorithm (K-means) with four risk factors (RA, Anti CCP, SJC, and ESR) to predict the occurrence of rheumatic illness. The research results show that any two of the four risk factors can diagnose rheumatoid arthritis [22]. Sheng et al. applied a Bayesian Network model to identify the risk of KOA in adults, the evaluation results show that the Bayesian model recorded a better identification result than the existing models [23]. Researchers in Jamshidi et al. and Du et al. documented the benefits of applying ML methods to KOA diagnoses [24, 25]. Du et al. applied ML techniques to predict KOA progression, Principal component analysis (PCA) was used to select relevant features of the KOA dataset, four (4) ML algorithms; Artificial Neural Network (ANN), SVM, Random Forest (RF), and NB, were used to build the progression prediction models. Kellogram and Lawrence (KL) grade and Joint Space Narrowing (JSM) were the two metrics used to measure the progression, ANN recorded the best KL grade prediction, while RF recorded the best JSN prediction grade [26].

Tiulpin et al. applied Convolution Neural Network (CNN) for KOA's automatic diagnosis from plain radiographs [15], the study recorded a quadratic Kappa Coefficient of 0.83 and average multi-class diagnosis accuracy of 66.71%. According to Jessica, the Magnetic Resonance Imaging (MRI) images and data of knee scans of 86 patients, which do not have KOA symptoms, were used to train Transport-based Morphometry, a new customized ML technique, to differentiate between patients who will progress or not progress to Osteoarthritis [27]. The model recorded a 78% accuracy in future Osteoarthritis cases. This result suggested the possibility of Osteoarthritis detection at a potentially reversible stage. Onan presented a comparative evaluation of six ensemble learning methods; (Bagging, Dagging, Multi Boost, Ada Boost, Random subspace, and Decorate) to improve the diagnosis predictions of fourteen breast cancer base models, the result of this work affirms the ability of the ensemble learning to improve the performances of base models in the medical domain [28]. Oguntimilehin et al. evaluated the performance of two stacked generalization meta-learning algorithms (Random Forest and NNGE), used to combine and improve the diagnosis of malaria using six machine learning base algorithms (PART, REP Tree, J48, Random Tree, RIDOR, and JRIP) [29]. The comparison of the obtained results in terms of prediction accuracy shows that NNGE, as a Meta learner, performs better than RF. Research in Olasehinde and Olayemi applied Multiple Model Tree (MMT) Meta algorithms and three base algorithms; (NB, KNN, and DT) to implement a stacked ensemble model for the improvement of diagnoses of Lower Respiratory Tract Infections in pediatric patients. The result shows that the MMT model presented the highest diagnosis precision improvement with the KNN models; 12.80% for the consistency feature model, 13.52% for the correlation feature model, and 12.37% for information Gain 18.35% for the whole feature model [16].

Stefanus et al. (2019) developed a deep learning ensemble model from features extracted from the X-ray and Canny edge detected images to detect Tuberculosis (TB). The proposed ensemble model was evaluated using two publicly available

datasets. The best result of 89.77% accuracy achieved by the proposed model shows that using different features extracted from different images can improve TB's detection rate. Verma and Pal applied three (3) ensemble techniques, Bagging, Adaboost, and gradient boosting, to enhance and improve the performances of six ML models used in the diagnoses and prediction of various classes of skin diseases. Evaluation of the ensemble models on the dermatology dataset shows an improved diagnostic accuracy and effectiveness of skin disease predictions. Khalid applied an ensemble model to improve the classification accuracy of heart disease, the predictions of three base models of LR, NB, and Multilayer perceptron were combined using a trial-and-error majority voting. The proposed ensemble model achieved a better classification accuracy of 88.8%, higher than the best of the three base models [21].

Though ensemble learning has been applied to improve the diagnostic accuracy of several diseases and infections. This is a novel study that applied ensemble learning to enhance the clinical diagnosis of KOA. The motivation for this work originated from the need to enhance the accuracy of the clinical diagnostic KOA.

3. Data Description and Modelling

Following the identification of KOA's risk factors from the literature review and the expert medical physicians, information on patients was collected from their case notes. The information contains the risk factors for the treatment of KOA at the Federal Medical Center Ido Ekiti, Nigeria. The dataset was extracted from the case notes (records) of 2237 patients, each of the records contained fourteen independent variables and one dependent variable. All the fourteen independent risks factors were validated by medical experts as one of or set of possible factors for identifying risks of Knee Osteoarthritis (dependent variable). The dataset was randomly split into training, and testing datasets in the ratio of 6 to 4, as shown in Table 1. The training and testing dataset comprised 1130 and 782 patients with KOA, respectively, and 212 and 113 patients did not have KOA. These patients who were not having KOA were diagnosed with other forms of arthritis and bones disorder such as Rheumatoid Arthritis, Inflammatory arthritis, Psoriatic, Arthritis, Osteoporosis, and Osteoporosis. The training dataset is an imbalanced dataset of a ratio of 6 to 1 of the infected patients to non-infected patients.

Table 1. *Distribution of KOA Dataset Splits.*

	Training Dataset	Testing Dataset	Total
KOA Infected	1130	782	1912
Non KOA Infected	212	113	325
Total	1342 (60%)	895 (40%)	2237

3.1. Clinical Dataset Pre-processing

Table 2 reported the description of the attributes of the KOA dataset, and its possible values as follows.

a) Gender is either male or female, KOA is more common in women than males [2].
b) Age; the ages of the patients included in the study range from twenty-one (21) years to eighty-five years (85). Age is a major risk factor of KOA [32].
c) Family History: This attribute is meant to find out about any history of KOA in the family of the patient, it is either yes or no.
d) Waist Hip ratio (WHR): This is the ratio of the waist to the hips, it is a good measure of fat distribution in the patient that assists in determining the patient's overall well-being. WHR of 0.80 or lower in women and 0.95 or lower in men are classified as low, WHR of between 0.81 to 0.85 in women and 0.96 to 1.0 in men are classified as moderate, while 0.86 or higher in women and above 1.0 in men are classified as high. According to Gandhi et al, high WHR increases the risk for heart disease and other conditions that are linked to being overweight, a healthy WHR is 0.9 or less in men and 0.85 or less in women [33].
e) Body Mass Index (BMI): This is the ratio of the height of the patient to weight, it is the ratio of patient's weight in kilograms to the square of his height in meters, and it is used to determine if the patient is obsessed or not. There are clinical shreds of evidence that demonstrate the existing relationship between the risk of knee KOA and BMI [34]. BMI of 18.5 is considered underweight, 18.5 to 24.9 is considered normal, 25.0 to 29.9 is considered overweight and above 30 is considered obsessed.
f) Hypertensive Heart Disease (HHD): This attribute whether the patient is having high blood pressure or not, HHD higher than 120/80 mmHg is considered high, it is either yes or no.
g) Joint pains: These are aches and discomfort experienced in any part of the body joints, it is either yes or no.
h) Cellulitis: This is a painful skin problem, it looks swollen and appears warm and red, it is yes when present and no when absent.
i) Seizure disorder: This is an abnormal electric activity in the brain, which may or may not cause dramatic noticeable symptoms, it is yes when present and no when absent.
j) Repeated stress on the Joint: This attribute is used to find out if there is repeated stress or pain on the patient's joint(s). Several studies have shown that weight reduction reduces KOA pains [31]. It is either yes or no, yes if repeated stress is experienced, and no if otherwise.
k) Limbs (legs) ulcer: This is a chronic long time wound in the leg that fails to heal after about 3 months or more of appropriate treatment, it is either yes or no.
l) Septic arthritis: This is a joint infection caused by germs, it is yes when present and no when absent.

m) Bone deformity: This is a distortion or movement of bone from its normal position in the body, often caused by diseases and bone injuries, it is yes when present and no when absent.
n) Joint Injuries: Strains, Sprains, Fractures, and Dislocation are four examples of joint injuries, it is either yes or no.
o) Class label: This indicate the diagnosis result of the patient, as either diagnosed with KOA or not.

All the attributes of the dataset were discretized to make it suitable for the ML model diagnosis building and evaluation, as reported in Table 2. All attributes with value "yes" were discretized as nominal value "one" (1), and attributes with value "no" were discretized as nominal value "zero" (0). Ages more than 20 years were discretized as ratio value "one" (1) and ages less than 20 were discretized as ratio value "zero" (0). The low WHR ratio is discretized as ordinal values "one" (1), moderate and high waist-hip ratios were discretized as ordinal values "two" (2) and "three" (3) respectively. BMI values; underweight, normal, overweight, and obsessed were discretized as ordinal values, "one" (1), "two" (2), "three" (3), and "four" (4) respectively.

Table 2. *Distribution of Identified Features in the Original Dataset.*

	Attribute Values	Attribute Type	Discretization of attribute values
Gender	Male, Female	Nominal	1 = male, 2= female
Age (years)	Above 21 to 85 years	Ratio	1 = ages > 20 years, 0 = ages < 20 years
Family History	Yes, No	Nominal	1 =Yes, 0 = no
Waist Hip Ratio	Low, Moderate, High	Ordinal	1 = low, 2= moderate, 3 = high
BMI	Underweight, Normal, overweight, Obsessed	Ordinal	1=Underweight, 2 = Normal, 3 = overweight, 4 = Obsessed
HHD	Yes, No	Nominal	1 =Yes, 0 = No
Joint pains	Yes, No	Nominal	1=Yes, 0 = No
Cellulitis of Leg	Yes, No	Nominal	1=Yes, 0 = No
Seizure Disorder	Yes, No	Nominal	1 =Yes, 0 = No
Ulcer of L/R Limb	Yes, No	Nominal	1 =Yes, 0 = no
Septic Arthritis	Yes, No	Nominal	1=Yes, 0 = No
Repeated stress on Joint	Yes, No	Nominal	1=Yes, 0 = No
Bone Deformities	Yes, No	Nominal	1 =Yes, 0 = No
Joint Injuries	Yes, No	Nominal	1=Yes, 0 = No
Class Label	Yes, No	Nominal	=Yes, 0 = No

3.2. Modelling Methodology

Modeling a successful and improved diagnosis model depends on the machine learning algorithm used in the model building and its eminence. This study applied an ensemble learning technique to optimize the diagnoses of KOA in adults. It employs five efficient and widely used ML base algorithms; LR, SVM, C4.5 DT, NB, and KNN for base-level diagnosis of KOA. Stacking with Multi Response Linear Regression (MLR) and Majority Voting were used for the Meta level combining the base models' diagnoses. The proposed system architecture is presented in Figure 2. It comprises three stages; the first stage involves the selection of attributes (risk factors) of the KOA dataset that are relevant to the diagnosis of the risk of KOA and the random split of the dataset into training and testing datasets in the ratio of 6 to 4. The second stage is the models' building stage, the KOA's training dataset was used to train and evaluate the five ML learning algorithms via ten folds cross-validation technique to ensure the dataset's reputation, in ten folds cross-validation, the whole training dataset is divided into ten folds, each of the ten folds is holds out for diagnoses evaluation while the rest nine folds are used for training in turns, the evaluated prediction of the base models was used to train the meta algorithms, and build the meta classifiers. During this stage, the base models and the ensemble models were built from the KOA training dataset. In the last stage, the base and the ensemble models built in the first stage were evaluated with the test dataset. The evaluation procedure is indicated in short dashed red lines in the figure. The base models were used to diagnose the test dataset; their diagnostic predictions were used to evaluate the Meta classifiers to obtain the final optimal diagnosis results. The base and ensemble models' comparative performance was measured using the evaluation metrics presented in section 3.5. The system was implemented using Python programming language on a Corel i3, 64bits, 2.4 GHz processor, 16MB Cache, 512GB SDD, Ms. Windows 7 operating system.

3.2.1. Feature Selection

The correlation filter-based feature selection technique was employed in this research to select the relevant features of the KOA dataset used to build diagnosis models with optimal diagnosis accuracy. Correlation Feature selection techniques generate all possible subsets of the KOA dataset and applied the merit function shown in equation (1) on all the subsets to determine the subset that is mostly correlated with the class label, the subset with the highest merit value is selected and returned as the selected relevant features of the KOA dataset.

$$M_s = \frac{k\,\bar{r}_{cf}}{\sqrt{k+k(k-1)\bar{r}_{ff}}} \quad (1)$$

where $\bar{r}_{cf}$ is the average class label to features, $\bar{r}_{ff}$ is the average features to features correlations and k is the number of features in the subset S.

3.2.2. Base Models

Five (5) Machine Learning Algorithms, LR, SVM, C4.5 DT, NB, and KNN, were adapted to build the base models.

(i). Logistic Regression

Logistic Regression (LR) is a simple supervised ML algorithm that models a relationship between independent features and the dependent response feature. It uses the logit function to predict the probability occurrences of binary classification of an event. LR assumes and treats all features (risk factor of OA) as independent of one another. Logistic regression is a model of choice for several medical data classification problems [29]. The logistic regression model is given in Equation (2)

$$In\left(\frac{p}{1-p}\right) = a_0 + a_1x_1 + a_2x_2 + \cdots \ldots \ldots . a_nx_n \qquad (2)$$

(ii). K-Nearest Neighbors (KNN)

KNN is a distance-based classification model capable of handling both binary and multi-class label classification. It is an instance-based learner that does less work during the training and more work during classification and prediction. Model evaluation with KNN is very computational and expensive. A new instance of a patient to be classified is compared against all instances in the KOA training dataset based on their Euclidean distance in equation (3). The label of the majority closest neighbor is returned as the label of the instance being classified. KNN was chosen as one of the base models in this research because of its ability to achieve high diagnostic accuracy in the domain of diagnosis of diseases as reported in [35].

$$d(p_i, q_t) = \sqrt{\sum_{i=1}^{n}(p_i - q_t)^2} \qquad (3)$$

(iii). Support Vector Machine

A support vector machine (SVM) is a powerful supervised learning algorithm used for analyzing data and pattern recognition. It is suitable for classification and regression problems based on a principle similar to KNN in that it represents the training set as points in an N-dimensional space and then attempts to construct a hyperplane that will divide the space into particular class labels with a clear margin of error. SVM constructs a separating optimal hyperplane with the largest margin between the dataset. It splits the dataset into two vector sets to identify two different classes in n-dimensional space vector, such that the margin between the classes is maximized and the distance between the hyperplane points is minimized. The equation of the class of hyperplanes is given in equation (4)

$$(w.x) + b = 0 \qquad (4)$$

Where w is the weight vector ($w \in R^N$), x is the input vector, b is the bias ($b \in R$). Corresponding to the decision function

$$F(x) = \text{sign}\ ((w.x) + b). \qquad (5)$$

SVM's brilliant performance on the medical dataset informs its choice as one of the proposed base learners. SVM achieved the best diagnostic accuracy than the NB classifier and Radial Basis Function (RBF) network classifiers for the binary diagnosis of three medical datasets; Heart dataset, Breast cancer dataset, and Diabetes dataset [36].

(iv). C4.5 Decision Tree

C4.5 Decision Tree classification model consists of nodes that the attribute names of the KOA dataset and the arcs that attribute (values) connected to other nodes to the leaves, which are the class label. A decision Tree (DT) builds a classification tree, uses it for the diagnosis, and predicts a new patient (either having KOA or not) in the KOA test dataset. DT calculates the Gain Ratio of all the training dataset attributes by dividing the attribute's information gain with its split value, Equation (6). The split value of an attribute is chosen by taking the average of all the values in the current attribute domain as given in Equation (7). The attribute with the highest gain ratio is chosen as the root attribute from all existing attributes. The root attribute divides the attributes into two branches. This procedure is repeated for each of the branches and all other subsequent branches until the tree is fully built. A new instance is classified as the tree node's leaf (class label) that satisfied its attributes' values.

$$Gain\ Ration\ A_i = \frac{Information\ Gain\ A_i}{Split\ Information\ A_i} \qquad (6)$$

$$Split\ info(A_i) = -\sum_{i=1}^{n}\frac{|t_j|}{|T|}\cdot \log_2\frac{|t_j|}{|T|} \qquad (7)$$

Where $|T|$ is the number of values of the current attribute, t is the values of attributes A_i, n is the number of values in attribute A_i.

The information gain of attribute X is given in Equation (8)

$$IG = H(Y) - H(Y|X) \equiv H(X) - H(X|Y) \qquad (8)$$

Where H(Y); the entropy of the attack categories and H(X|Y)entropy of class label given a certain attribute are is given by Equation (9) and (10)

$$H(Y) = -\sum_{c\epsilon C} p(y)\log_2\big(p(y)\big)\,; y\epsilon Y \qquad (9)$$

$$H(X|Y) = -\sum_{x\epsilon X} p(y)\sum_{c\epsilon C} p(y|x)\log_2\big(p(y|x)\big) \qquad (10)$$

Where;$y\epsilon Y\ and\ x\epsilon X$

C4.5 DT algorithm was chosen as one of the base algorithms in this research because of its good performance in [37].

(v). Naive Bayes

The work in [38] described Naïve Bayes algorithm to be good at solving diagnostic and predictive problems. NB is a probabilistic classifier. It assumes that the attributes of the KOA dataset are independent of each other, it calculates the probability for each symptom in the risk factor given a class label to obtain a joint conditional probability for each patient's risk factors and then use Bayes rule to derive conditional probability for each class label. A given risk factor is diagnosed as the class label with the highest probability value. The probability that a class label y_j will be assigned to a given unlabeled instance X of the KOA dataset is given in Equation (11).

$$p\big(y_j\,|x_1,,,x_{18}\big) = \frac{p(y_j)p(x_i|y_j)}{p(x_i)}\ (\forall_j = 0,1) \qquad (11)$$

The maximum posterior probability for classifying a new instance as a class label is given in Equation (12)

$$y = \underset{y}{\arg max}\, py_j \prod_{j=0}^{1} p\,(y_j) p\,(x_1, x_2, \ldots x_{18} \mid y_j) \quad (12)$$

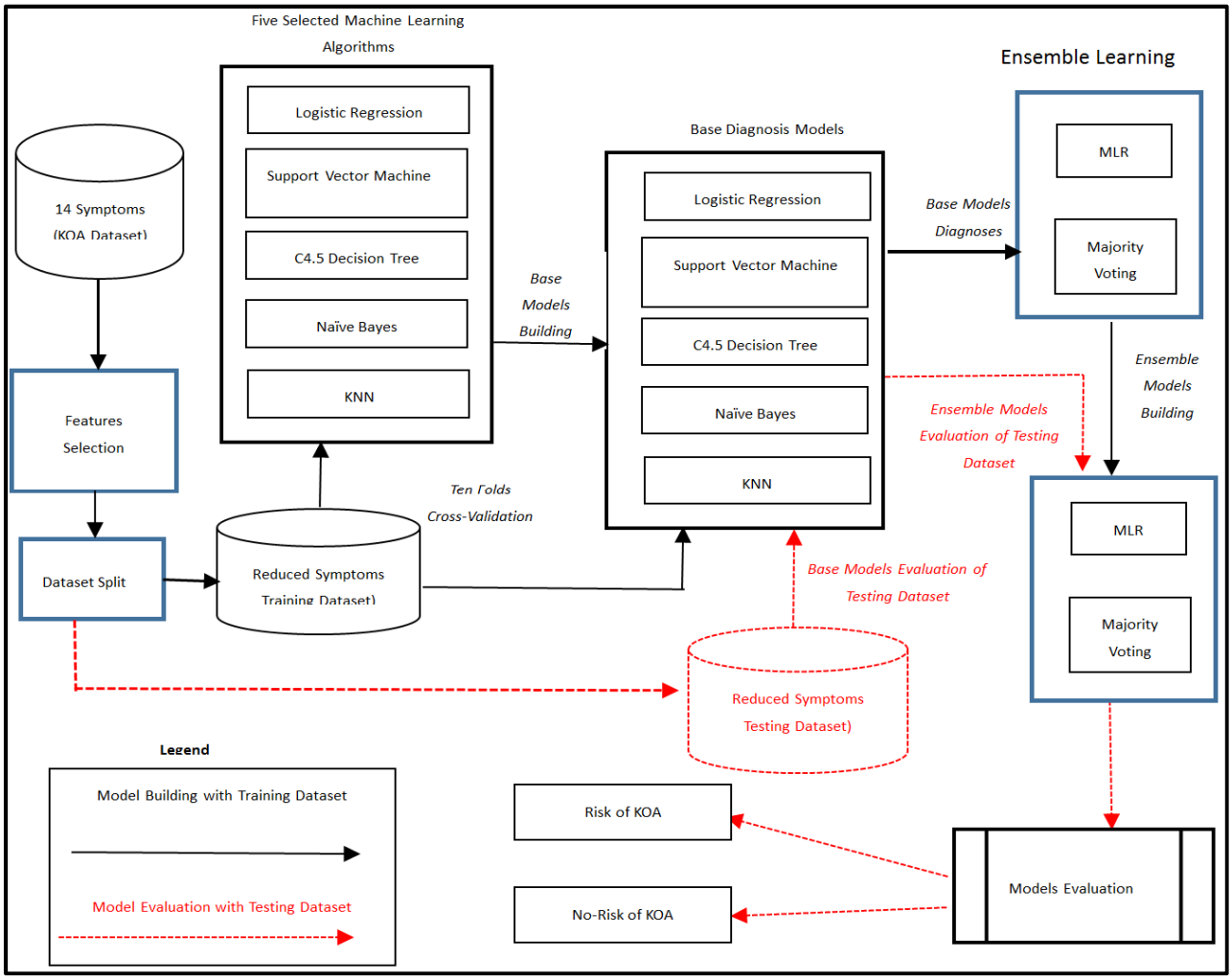

Figure 2. *Architecture of the Ensemble Methods Improvement of Knee Osteoarthritis Risk System.*

3.2.3. Ensemble Models

Ensemble learning combines individual predictions of base learners to improve the overall performance of the predictive model. Several studies have shown that stacked ensemble diagnoses of two or more diagnosis models using Meta classifiers improve their clinical diagnosis [39]. Meta Classifiers are supervised learning techniques that systematically learn several based learner models' predictions to build a meta-model to improve the base models' effectiveness. The meta-learning algorithm transforms an aspect of a learning algorithm such that the transformed learner is better than the original learner at learning from other experiences. In this research, two ensemble techniques; stacking with Multiple Linear Regression and Majority Voting, were individually used to combine and improve the five base models' diagnosis.

(i). Multi-Response Linear Regression

Multi-Response Linear Regression (MLR) Algorithm is a form of linear regression that models the relationship between two or more attributes known as independent variables and response attributes (class label) known as the dependent variable. MLR is an adaptation of multiple linear regression represented in equation (13)

$$Y = \beta_o + \beta_1\, \hat{y}_{i1} + \beta_2\, \hat{y}_{i2} + \ldots + \beta_n\, \hat{y}_{in}\ (i = 1,2,\ n=1,..3) \quad (13)$$

where Y is the final (combined) diagnosed value of the MLR stacked ensemble (dependent variable), $\hat{y}_1$, $\hat{y}_2$,...., $\hat{y}_n$ are the diagnoses of the base models (independent variables), n is the no of base models used, β_0 is the value of Y when all the independent attributes are zeroes, β_1, β_2...., β_n are the estimated coefficients of $\hat{y}_1$, $\hat{y}_2$,...., $\hat{y}_n$. For each target class C_m, a liner equation LR_m is constructed and the values of β_1, β_2...., β_n in the regression equations are computer as in equation (14).

$$\left.\begin{aligned} Y &= \beta_0 + \beta_1 \hat{y}_{1,1} + \beta_2 \hat{y}_{1,2} + \beta_3 \hat{y}_{1,3} \\ Y &= \beta_0 + \beta_1 \hat{y}_{2,1} + \beta_2 \hat{y}_{2,2} + \beta_3 \hat{y}_{2,3} \end{aligned}\right\} \quad (14)$$

The computed estimated values of β1, β2..., β_n will be substituted in equation (13) to generate the MLR equation, which is use to improve the base models diagnoses of the

KOA patient. A new Instance of the base models diagnoses ($\hat{y}_1$, $\hat{y}_2$... $\hat{y}_n$) will be substituted in the MLR equation and diagnosed as the class C_m with maximum value of $LR_m(Y)$.

(ii). Majority Voting (MV)

The majority voting (MV) rule is an Ensemble Learning technique in which every base model makes a diagnosis (vote) for each test instance and the final diagnosis is the diagnosis with a majority vote [40]. MV predicts and returns the most diagnosed class label by the base models. It is represented as

$$Y = \text{Mode } \{C_1(\hat{y}_1),\ C_2\ (\hat{y}_2),...,\ C_p(\hat{y}_p)\} \quad (15)$$

where C_1, C_2, C_p are the base models (classifiers), $\hat{y}_1$, $\hat{y}_2$,...., $\hat{y}_p$ are the prediction of the base models (classifiers), and p is the number of base models.

3.3. Statistical Test

Statistical tests have severally been used to validate and compare the results of computational intelligence models. Two statistical tests; Pairwise Differences Mean Ranking and Kendall's tau_b Correlations were used to validate and compare the diagnostic results of the five machine learning models and the two ensemble learning models.

3.3.1. Pairwise Difference Means

The pairwise difference means measures and ranks the error of deviation between the actual class labels and the class labels diagnosed by each of the computational intelligence models. Pairwise Mean Difference (PMD) is given in equation (16), the lower value PMD, the better the diagnostic performance of the model.

$$APMD = \frac{\sum_{i=1}^{n} |(x_i - y_i)|}{n} \quad (16)$$

Where n = no of observation, x_i *and* y_i *are* the actual and the diagnosed label of observation i respectively.

3.3.2. Kendall's tau_b Correlations

The Kendall tau-b measures the strength of association between variables X (Actual class label) of the KOA test dataset *and* Y (diagnosed class label) of the diagnostic model evaluation of the KOA test dataset, it is given in equation (17). The values of Kendall's tau_b range between -1 to +1, the closer the value to +1 the stronger the correlation between the actual class label and the diagnosed class label.

$$\text{Kendall's } Taub = \frac{P-Q}{\sqrt{(P+Q+X_0)}\sqrt{(P+Q+Y_0)}} \quad (17)$$

where P is the number of concordant pairs, Q is the number of discordant pairs, X_O is the number of pairs tied only on the X variable, Y_0 is the number of pairs tied only on the Y variable.

3.4. Performance Metrics

The confusion matrix shows the distribution of instances that are either correctly classified or wrongly classified by the diagnosing models. It consists of four possible outcomes: True Positive (TP), the number of KOA patients diagnosed as having KOA. False Negative (FN) is the number of KOA patients that were diagnosed as not having KOA. In the same vein, False Positive (FP) is the number of patients that are not having KOA but were diagnosed as having KOA, False Negative (TN) is the number of patients that are not having KOA and were diagnosed as not having it. Accuracy and Error rate are the two most common intuitive metrics derivable from the confusion matrix. Accuracy is the ratio of all correct diagnoses to the total number of patients diagnosed, while the Error rate is the ratio of all incorrect diagnoses to the total number of patients diagnosed. The Formulae for Accuracy and Error rate are given in equations (18) and (19), respectively.

$$Accuracy = \frac{TP+TN}{TN+FP+FN+TP} \quad (18)$$

$$Error\ Rate = \frac{FP+FN}{TN+FP+FN+TP} \quad (19)$$

Other intuitive metrics of the confusion matrix used in the evaluation of the diagnostics performance of models used in this research are; Precision and sensitivity.

3.4.1. Precision

Precision is the ratio of true positives to the sum of true positives and false positives. High Precision implies a low false-positive rate. It is given in equation (20).

$$Precision = \frac{TP}{FP+TP} \quad (20)$$

3.4.2. Sensitivity/Recall

Sensitivity, also known as a recall, is the correct positive diagnosis ratio to the total number of actual positive instances. High sensitivity is desirable; it implies that infected patients are correctly diagnosed as having the disease. It is given in equation (21).

$$Sensitivity = \frac{TP}{FN+TP} \quad (21)$$

3.4.3. F1-Score

Accuracy metrics can be misleading when dealing with an imbalanced dataset. In such cases, other evaluation metrics should be considered in addition to accuracy. The F1-Score is a good metric for evaluating imbalanced data [41]. Authors in [42] reported that F1-Score gives a better model performance measurement than the accuracy metric evaluating imbalanced class distribution.; it is the harmonic average of sensitivity, and precision is given in equation (21), the values of F1 scores range between zero to one, F1 score of zero (0) value implies the diagnosis model is imperfect (not good) while F1 score of one (1) implies a very good and perfect model, the closer the value of F1 score to one (1), the better, the diagnosis of the model. A good F1 score indicates a lower or no misdiagnosis of both the infected and non-infected patients.

4. Experimental Results and Discussion

Table 3 reports the results of the correlation feature selection technique, seven features out of the fourteen features

of the KOA dataset were selected as the feature subset that is mostly correlated to the class label. Table 4 reports the base models' confusion matrix and how each of the base diagnosis models performed Figure 4 reported the diagnosis accuracy and diagnosis Error rate recorded by each of the base models. SVM recorded the highest correct diagnostic accuracy of 87.93%. LR, C4.5 DT, KNN, and NB recorded correct diagnostic accuracy of 85.70%, 85.14%, 84.47%, and 82.79% respectively. The SVM, LR, C4.5 DT, KNN, and NB diagnosis models respectively diagnosed 76.11%, 72.57%, 70.80%, 66.37%, and 64.60% of the 113 patients that are having other forms of bones disorder such as Rheumatoid Arthritis, Inflammatory arthritis, Psoriatic Arthritis, Osteoporosis, and Osteonecrosis were diagnosed as not having KOA infected. while the SVM, LR, C4.5 DT, KNN, and NB models diagnosed 23.89%, 27.43%, 29.20%, 33.63% and 35.40% of this categories respectively as being infected with KOA.

SVM model recorded the lowest diagnostic error rate of 12.07% and the highest true positive rate of 89.64%. It shows that the SVM model performs better than all the other base models. The lowest true positive value of 85.42% and the highest diagnostic error rate of 17.21% recorded by the NB model implies that the NB model recorded the least diagnosis performance. Out of the 895 patients who presented for diagnosis, SVM achieved the highest correct diagnosis of 786 patients with a misdiagnosis of 109 patients. LR achieved the correct diagnosis of 767 patients with a misdiagnosis of 128 patients. NB recorded the highest misdiagnosis of 155 patients.

Table 3. *Features of KOA dataset selected by Correlation Features Selection Technique.*

Selected Features
Waist Hip Ratio, Septic Arthritis, Seizure Disorder, Joint Pains, Joint Injuries, Joint stress, History

Table 4. *Confusion Matrix and Diagnosis Performances of Base Models.*

Base Models	TP	FN	TN	FP	Diagnosis Accuracy	Diagnosis Error Rate	Model Precision	True Positive Rate Model Sensitivity	False Positive Rate	F1 Score
SVM	701	81	86	27	87.93%	12.07%	96.29%	89.64%	23.89%	0.9285
LR	685	97	82	31	85.70%	14.30%	95.67%	87.60%	27.43%	0.9146
C4.5 DT	682	100	80	33	85.14%	14.86%	95.38%	87.21%	29.20%	0.9111
KNN	681	101	75	38	84.47%	15.53%	94.71%	87.08%	33.63%	0.9073
NB	668	114	73	40	82.79%	17.21%	94.35%	85.42%	35.40%	0.8966

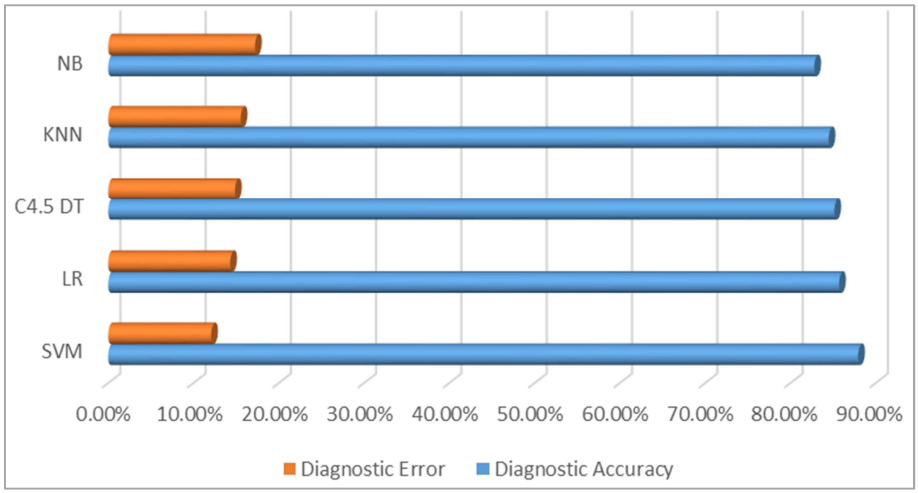

Figure 3. *Diagnosis Accuracy and Error Rate of the Base Diagnostic Mode.*

4.1. Evaluation of Machine Learning Models

Kokkotis et al. applied seven ML algorithms to implement seven KOA diagnostic models with the osteoarthritis initiative (OAI) database. Five out of these seven implemented models were also implemented as base models in the proposed models [43]. Table 5 shows the comparison between the proposed models and the work of Kokkotis et al. in terms of diagnostic accuracy. In each of the works under comparison, SVM models achieved the best diagnostic accuracy. The SVM achievement could be attributed to its ability to be well suited for diagnosing small and medium-size datasets regarding the number of instances and dimensionality. LR achieved the second-best accuracy in both compared models. C4.5 DT achieved the third-best accuracy in the proposed model and the least diagnostic accuracy in [43]; the reason for this disparity could be attributed to the strength of different DT algorithms, while the proposed model implemented C4.5 DT, the type of DT model implemented in the Kokkotis et al. was not mentioned. The performance of NB is closely higher than that of KNN in both research works. All the five base models of the proposed model achieved higher diagnostic accuracy than their corresponding models in [43].

Table 5. *Comparison of the Proposed Base Models and Other states of the art performance.*

Proposed Base Models Result		**[43], Kokkotis et al. Result**	
Base Models	**Diagnostic Accuracy**	**Machine Learning Models**	**Diagnostic Accuracy**
SVM	87.93%	SVM	74.07%
LR	85.70%	LR	72.84%
C4.5 DT	85.14%	DT	61.73%
KNN	84.47%	KNN	71.60%
NB	82.79%	NB	68.52%
		Random Forest	67.11%
		xgboost	70.47%

Table 6 reports the confusion matrix of the ensemble models and their diagnostic accuracy and Error rate. MLR stacked ensemble recorded the best ensemble diagnosis accuracy of 97.77%; Majority Voting recorded 96.54% accuracy. Both ensemble models perform better than each base model, thus improving the base models' diagnostic accuracy and reducing their error diagnosis rate. The ensemble models reduced the lowest misdiagnoses of the 109 patients recorded by SVM and the highest misdiagnosed 155 patients recorded by NB to 20 misdiagnosed patients by MLR and 31 misdiagnosed patients by MV. Figure 4 reports the diagnostic accuracy and Error rate recorded by the base and the ensemble models. Figure 5 reports the evaluation of the Base and Ensemble Models based on their F1 score values. F1 score is considered a better metric than accuracy for evaluating the ML models' performance trained with an imbalance dataset [44]. According to Jack (2020), accuracy provides an over-optimistic of the model toward the majority class; hence it does not always give a full picture of the model performance. The closer the F1 score of a model to one (1), the better the model [45]. SVM model with an F1 score of 0.9285 is the best model among the base models, NB being the model with the least performance achieved the least F1 score of 0.8966. The ensemble models achieved a higher F1 score than all the base models. The highest F1 score of 0.9872 achieved by MLR models is slightly higher than the 0.9801 recorded by the MV.

Table 6. *Confusion Matrix and Diagnosis Performances of Ensemble Models.*

Base Models	**TP**	**FN**	**TN**	**FP**	**Diagnosis Accuracy**	**Diagnosis Error Rate**	**Model Precision**	**Model Sensitivity**	**False Positive Rate**	**F1 Score**
MLR	774	8	101	12	97.77%	2.23%	98.47%	98.98%	10.62%	0.9872
VOTING	766	16	98	15	96.54%	3.46%	98.08%	97.95%	13.27%	0.9801

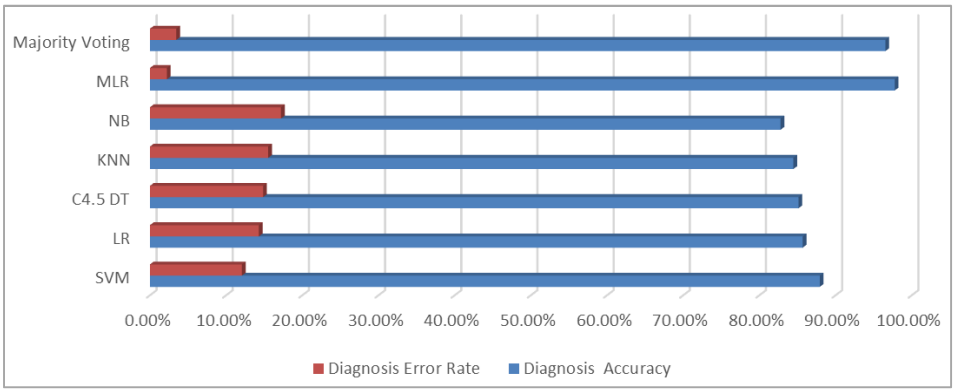

Figure 4. *Accuracy and Error Rate obtained by all the Diagnoses Models.*

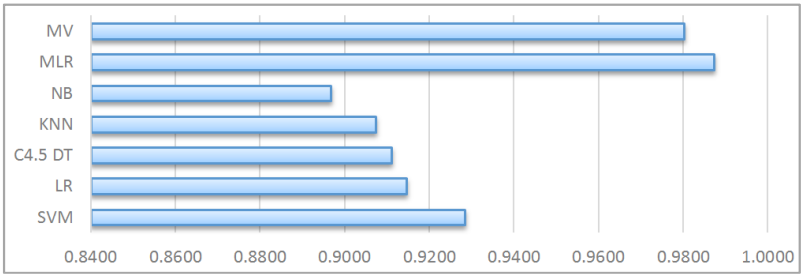

Figure 5. *F1 Scores of the Base and Ensemble Models Diagnoses.*

Tables 7 and 8 report the ensembles' diagnosis accuracy improvements of the base models. Both ensemble models recorded improved diagnosis accuracies on the base models. MLR stacked ensemble recorded the greatest improvement of

18% with the NB base model and the Lowest improvement of 11.19% with the SVM base model. The majority Voting ensemble recorded the lowest improvement of 9.79% with the SVM base model and the greatest improvement of 16.61% with the NB base model. MLR achieved higher and better diagnostic accuracy improvement than MV. Figure 6 reports the graphical comparison of the diagnostic accuracy improvement of the two ensemble models.

Table 7. *MLR Ensemble Models Diagnosis Accuracy Improvement of Base Models.*

Base Models	Diagnosis Accuracy	MLR Ensemble Diagnosis Accuracy	MLR Diagnosis Accuracy Improvement
SVM	87.93%		11.19%
LR	85.70%		14.08%
C4.5 DT	85.14%	97.77%	14.83%
KNN	84.47%		15.75%
NB	82.79%		18.09%

Table 8. *Majority Voting Ensemble Models Diagnosis Accuracy Improvement of Base Models.*

Base Models	Diagnosis Accuracy	Voting Ensemble Diagnosis Accuracy	Majority Voting Diagnosis Accuracy Improvement
SVM	87.93%		9.79%
LR	85.70%		12.65%
C4.5 DT	85.14%	96.54%	13.39%
KNN	84.47%		14.29%
NB	82.79%		16.61%

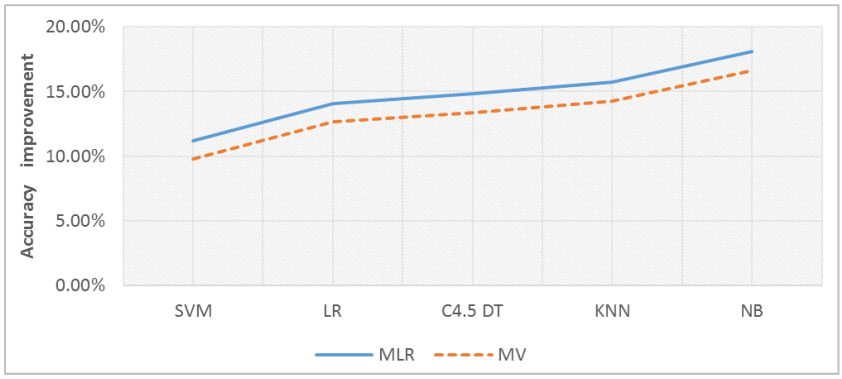

Figure 6. *Line Graph of Diagnosis Accuracy Improvement by the Ensemble Models.*

4.2. Validation of Research Results

Tables 9 and 10 show the results of the two statistical tests carried out to validate the computational intelligence results of the five base and the two ensemble diagnostic models. The two techniques gave the same results based on their ranking, the pairwise difference means result presented in table 8 reported MLR to have the least error of deviation between the actual and diagnosed class label. From table 9, Kendall's tau_b result of MLR shows the strongest correlation between the actual and diagnosed class label. The results of the statistical tests confirm the improved diagnostic performance of the two ensemble models.

5. Conclusion

KOA is the most common form of arthritis and one of the leading causes of disability globally, affecting 3.8% of the global population. Despite being one of the leading causes of disability worldwide, the pathophysiology of this disease is unknown, and the single most effective involvement in treating the symptoms of knee OA is not clear. Knee replacement surgery remains the only effective cure for KOA at its advanced stage. Early diagnosis of KOA has proven to ensure its proper management, prolong healthy patient-years, prevent cartilage from falling apart to slow down its progression and reduce the effect of its future disability. Machine and Ensemble learning has been applied to improve the diagnostic accuracy of several diseases and infections. Ensemble learning has not been applied to improve KOA's clinical diagnostic accuracy, to the best of our knowledge; this is the first work that applied ensemble learning to improve the clinical diagnostic accuracy of the risk of KOA. This study examines the strength of two ensemble learning methods to optimize the clinical diagnostic accuracy of KOA in Adults. The patient's clinical information dataset used in this research was obtained from FMC, Ido Ekiti. Nigeria. The results of this study established the possibility of improving the clinical diagnosis of the risk of KOA using ensemble learning methods. The comparison of the clinical diagnoses of the base models with a current state of arts similar study confirmed the superiority of this study in terms of diagnostic accuracy and performance. The performance of stacking with MLR was slightly higher than the majority voting method. Statistical analysis of the actual patient's status and the diagnosed results confirms the accurate performance of the base and ensemble models. For future work, the authors seek to apply ML and EL

to analyze the CT scan images of the Knees of younger people for early prediction of the risk of KOA disease in the future.

Table 9. *Paired Differences Mean (Diagnostic Error Measurement) of the Actual and Diagnosed class label.*

		Absolute Paired Mean Difference	Std. Deviation	RANK
Pair 1	Class Label - SVM	0.080	0.345	3
Pair 2	Class Label – LR	0.097	0.361	4
Pair 3	Class Label - C4.5	0.106	0.377	5
Pair 4	Class Label – KNN	0.101	0.377	6
Pair 5	Class Label – NB	0.127	0.392	7
Pair 6	Class Label – MV	0.004	0.134	2
Pair 7	Class Label – MLR	0.003	0.129	1

Table 10. *Kendall's tau_b Correlations of the Actual and Diagnosed class label.*

	Class Label	SVM	LR	C4.5	KNN	NB	MV	MLR	RANK
Class Label	1.000								
SVM	0.591**	1.000							3
LR	0.568**	0.486**	1.000						4
C4.5	0.537**	0.482**	0.502**	1.000					5
KNN	0.531**	0.454**	0.451**	0.434**	1.000				6
NB	0.518**	0.505**	0.456**	0.545**	0.485**	1.000			7
MV	0.922**	0.656**	0.615**	0.599**	0.577**	0.564**	1.000		2
MLR	0.927**	0.652**	0.611**	0.595**	0.581**	0.567**	0.995**	1.000	1

**Correlation is significant at the 0.01 level (2-tailed)

6. Recommendation

Due to the high diagnostics accuracy rate recorded by the base models and its improvement by the ensemble models, the hybridized system from this research is highly recommended for the diagnosis and early detection of the risk of KOA in Adults, this enables early intervention of the pharmacological management of its risk.

Conflict of Interest

The authors stated no conflict of interest; this manuscript is not under consideration or review with any journal.

Acknowledgements

The authors acknowledged the support and contributions of the Ethical Committee and staff of the record department, Federal Medical Centre, Ido Ekiti, Nigeria, for providing the dataset used in this research.

References

[1] Ayhan E., Kesmezacar H. & Akgun I. 2014. Intra-articular injections (corticosteroid, hyaluronic acid, platelet-rich plasma) for knee osteoarthritis. World J Orthop. 5 (3), 351–361.

[2] Cross M., Smith E., Hoy D., Nolte S., Ackerman I. & Fransen M. 2014. The Global Burden of Hip and knee Osteoarthritis: Estimates from the Global Burden of Disease 2010 study. Ann Rheum Dis, (73): 1323-1330.

[3] Heidera B., 2011. Knee Osteoarthritis Prevalence, Risk Factors, Pathogenesis, and Features: Part I. Caspian J Intern Med (2): 205-12, 2011.

[4] Palazzo C., Rayaud J., Papelard A., Rayau P. & Poirauden D. S. 2014. The Burden of Musculoskeletal Conditions. Plos ONE. 9 (3: e90633.

[5] Ogunlade, S. O., Alonge, T. O., Omololu, A. B. & Adekolujo, O. S. 2005. Clinical spectrum of large joint osteoarthritis in Ibadan, Nigeria. European J. Science Res. 11: 116-122.

[6] Palazzo C., Nguyen C., M. Lefevre-Colau M., Rannou F. & Poiraudeau S. 2016. Risk Factors and Burden of Osteoarthritis. Annals of Physical and Rehabilitation Medicine. 59 (3): 134-138, DOI: 10.1016/j.rehab.2016.01. 006.

[7] Christensen R., Bartels E. M. & Bliddal A. 2007. Effect of Weight Reduction in Obese Patients Diagnosed with Knee Osteoarthritis: A Systematic Review and Meta-Analysis. Ann Rheum Dis, (66): 433-439.

[8] Zheng H. & Chen C., 2015. Body Mass Index and Risk of Knee Osteoarthritis: Systematic Review and Meta-Analysis of Prospective Studies. BMJ Open.; 5 (12): e007568. DOI: 10.1136/BMJ open-2014-007568.

[9] Murphy L., Schwartz T. A., Helmick C. G, Renner J. B., Tudor G. & Koch G., Dragomir A., Kalsbeek W. D., Luta G. & Jordan J. M. 2008. The lifetime risk of symptomatic knee Osteoarthritis. Arthritis and rheumatism. 59 (9): 1207–1213. https://doi.org/10.1002/art.24021

[10] Dulay GS, Cooper C, Dennison EM. Knee pain, knee injury, knee osteoarthritis and work. Best Pract Res Clin Rheumatol. 2015; 29 (3): 454–461.

[11] Web1. https://www.mayoclinic.org/diseases-conditions/ Osteoarthritis /symptoms-causes/syc- 20351925 (Accessed 17th October, 2019).

[12] Sharma V., Anuvat K., John L., Davis M. 2017. Scientific American Pain Management - Arthritis of the knee. Decker: Pain-related disease states.

[13] Esser S., Bailey A., 2011. Effects of exercise and physical activity on knee osteoarthritis. Curr Pain Headache Rep. 15 (6): 423–430.

[14] Shinjini K., Beth G. A., Mustapha B., Erik B. D., Shadpour D., Mohammad S. R., Richard G. S., Kenneth L. U. & Gustavo K. R. 2020. Enabling early detection of Osteoarthritis from presymptomatic cartilage texture maps via transport-based learning. Proceedings of the National Academy of Sciences Oct 2020, 117 (40): 24709-24719; DOI: 10.1073/pnas.1917405117. DOI: 10.1109/ICTCS.2019.8923053.

[15] Tiulpin A., Thevenot J., Rahtu E., Lehenkari P. & Saarakkala. 2018. Automatic Knee Osteoarthritis Diagnosis from Plain Radiographs: A Deep Learning-Based Approach. Sci Rep 8, 1727. https://doi.org/10.1038/s41598-018-20132-7

[16] Olasehinde O. O. & Olayemi O. C. 2019. Stacked Ensemble Approach to the Development of Lower Respiratory Tract Infection Diagnoses System International Journal of Computer Science and Network, 8 (5): 421-435, ISSN (Online): 2277-5420. www.IJCSN.org

[17] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter. S. M.., Blau, H. M., & Thrun, S. 2017. Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542: 115-118.

[18] Olasehinde O. O., Williams K. & Olayemi O. C. 2018. A Machine Learning Framework for Improving Classification Accuracy Using Stacked Ensemble, Proceedings of the 14th iSTEAMS Multidisciplinary Conference, AlHikmah University, Ilorin, Kwara State, Nigeria. 117-124.

[19] Verma, A. K. & Pal, S. 2020. Prediction of Skin Disease with Three Different Feature Selection Techniques Using Stacking Ensemble Method. Appl Biochem Biotechnol 191, 637–656. https://doi.org/10.1007/s12010-019-03222-8

[20] Rajaraman S., Candemir S., Alderson P. O., Xue Z., Kohli M., Abuya J., Thoma G. R. & Antani S. 2018. "A novel stacked generalization of models for improved TB detection in chest radiographs," 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, 2018, pp. 718-721, DOI: 10.1109/EMBC.2018.8512337.

[21] Khalid R., 2019. Improving the prediction accuracy of heart disease with ensemble learning and majority voting rule. Chapter 8 in U-Healthcare Monitoring Systems. pp. 179-196. DOI: https://doi.org/10.1016/B978-0-12-815370-3.00008-6.

[22] Yoo J., Lim M. K., Ihm C., Choi E. S. & Kang M. S. 2017. A Study on the Prediction of Rheumatoid Arthritis using Machine Learning. Int. Journal of Applied Engineering Research 12 (20): 9858–9862, 2017. http://www.ripublication.com.

[23] Sheng B., Huang L., Wang X., Zhuang J., Tang L., Deng C., & Zhang Y. 2019. Identification of Knee Osteoarthritis Based on Bayesian Network: Pilot Study'' JMIR Med Inform; 7 (3): e13562. URL: http://medinform.jmir.org/2019/3/e13562/ DOI: 10.2196/13562 PMID: 31322132.

[24] Jamshidi A., Pelletier J., Martel-Pelletier J. 2019. Machine-Learning-Based Patient-Specific Prediction. Models for Knee Osteoarthritis. Nat. RevRheumatol 15: 49–60. https://doi.org/10.1038/s41584-018-0130-5.

[25] Kluzek S., and Mattei T. A., 2019. Machine-learning for osteoarthritis research. Osteoarthritis Cartilage, 27 (7): 977-978.

[26] Du Y., Almajalid R., Shan J. & Zhang M. 2018. A Novel Method to Predict Knee Osteoarthritis Progression on MRI Using Machine Learning Methods. IEEE Trans Nanobioscience. 17 (3): 228-236, DOI: 10.1109/TNB.2018.2840082.

[27] Jessica K., 2020. Artificial Intelligence May Predict Osteoarthritis Years before Onset. (Accessed 27th November 2020) https://healthitanalytics.com/news/artificial-intelligence-may-predictsteoarthritis- years-before-onset.

[28] Onan A. (2015) On the Performance of Ensemble Learning for Automated Diagnosis of Breast Cancer. In: Silhavy R., Senkerik R., Oplatkova Z., Prokopova Z., Silhavy P. (eds) Artificial Intelligence Perspectives and Applications. Advances in Intelligent Systems and Computing, vol 347. Springer, Cham. https://doi.org/10.1007/978-3-319-18476-0_13

[29] Oguntimilehin A., Adetunmbi O., & Osho I. 2019. Towards Achieving Optimal Performance using Stacked Generalization Algorithm: A Case Study of Clinical Diagnosis of Malaria Fever. The International Arab Journal of Information Technology, (16) 6.

[30] Stefanus K. TH, Mohd H. A. H., Abdullah B., Razali Y., Mohammad S. J. 2019. Ensemble deep learning for tuberculosis detection using chest X-Ray and canny edge detected images. IAES International Journal of Artificial Intelligence (IJ-AI) (8) 4: 429-435. ISSN: 2252-8938, DOI: 10.11591/ijai.v8.i4.pp429-435.

[31] Christensen R, Bartels EM, Astrup A*, et al, 2007.* Effect of weight reduction in obese patients diagnosed with knee osteoarthritis: a systematic review and meta-analysis. *Annals of the Rheumatic Diseases.* (66): 433-439.

[32] Felson D. T., Lawrence R. C., Dieppe P. A., Hirsch R., Helmick C. G., and Jordan J. M.: Osteoarthritis: new Insights. Part 1: The Disease and its Risk Factors. Ann Intern Med, Vol. 133, 635-646, (2000).

[33] Gandhi R, Dhotar H, Tsvetkov D, Mahomed NN. The relation between body mass index and waist-hip ratio in knee osteoarthritis. Can J Surg. 2010 Jun; 53 (3): 151-4. PMID: 20507785; PMCID: PMC2878991.

[34] Richette P., Poitou C., Garnero P., Vicaut E., Bouillot J. L., and Lacorte J. M.: Benefits of Massive Weight Loss on Symptoms, Systemic Inflammation, and Cartilage Turnover in Obese Patients with knee Osteoarthritis. Ann Rheum Dis. Vol. 70, No. 1, 139-144, (2011), DOI: 10.1136/ard.2010.134015.

[35] Dudoit S., Friday J., & Speed T. P. 2002. Comparison of Discrimination Methods for the Classification of Tumors Using Gene Expression Data". J Am Stat Assoc, 97: 77–87.

[36] Janardhanan P., Heena L. & Sabika F. 2015. Effectiveness of Support Vector Machines in Medical Data Mining. Journal of communications software and systems, 11: 25-30.

[37] Wharton W., Kusnanto H. & Herianto H. 2016. Interpretation of Clinical Data Based on C4.5 Algorithm for the Diagnosis of Coronary Heart Disease. Healthcare informatics research, 22 (3), 186–195. https://doi.org/10.4258/hir.2016.22.3.186

[38] Taheri S. O. N. A., 2015. Learning the Naive Bayes Classifier with Optimization Model. 23 (4): 787–795. DOI: 10.2478/amcs-2013-0059.

[39] Srimani1 P. K. and Manjula S. K. 2013. Medical Diagnosis Using Ensemble Classifiers - A Novel Machine-Learning Approach, Journal of Advanced Computing 1 (9-27) doi: 10.7726/jac.2013.1002

[40] Atallah R. & Al-Mousa A. 2019. Heart Disease Detection Using Machine Learning, Majority Voting Ensemble Method. 2nd International Conference on New Trends in Computing Sciences (ICTCS), Amman, Jordan, 1-6.

[41] Jeni L. A, Cohn J. F., De La Torre F. (2013). Facing imbalanced data– recommendations for the use of performance metrics, in 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, IEEE, 2013, pp. 245–251.

[42] Siblini W., Fréry J., He-Guelton L., Oblé F., Wang YQ. (2020) Master Your Metrics with Calibration. In: Berthold M., Feelders A., Krempl G. (eds) Advances in Intelligent Data Analysis XVIII. IDA 2020. Lecture Notes in Computer Science, vol 12080. Springer, Cham. https://doi.org/10.1007/978-3-030-44584-3_36

[43] Kokkotis C., Moustakidis S., Giakas, G., Tsaopoulos, D. 2020. Identification of Risk Factors and Machine Learning-Based Prediction Models for Knee Osteoarthritis Patients. Appl. Sci. 10, 6797.

[44] Akosa, J., 2017. Predictive Accuracy: A Misleading Performance Measure for Highly Imbalanced Data. Proceeding of A New Era of SAS Global AI and Analytics Events, session 0942-2017. https://support.sas.com/resources/papers/proceedings17/0942-2017.pdf

[45] Jack T. (2020, December 2), Beyond Accuracy: other Classification Metrics you should know in Machine Learning. https://towardsdatascience.com/beyond-accuracy-other-classification-metrics-you-should-know-in-machine-learning-ea671be83bb7