International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V73I4P131

Volume 73 Issue 4, 386-395, April 2025

Original Article

AI for Healthcare: Improving Classification Models for Screening Osteoarthritis Patients with Voting and Embedded Learning Techniques

Ploykwan Jedeejit[1], Wongpanya S. Nuankaew[2], Patchara Nasa-ngium[3], Pratya Nuankaew[4]

[1]*Department of Marketing, Faculty of Business Administration, Bangkok Thonburi University, Bangkok, Thailand.*
[2]*Department of Computer Science, School of Information and Communication Technology, University of Phayao, Phayao, Thailand.*
[3]*Department of Computer Science, Faculty of Science and Technology, Rajabhat Maha Sarakham University, Maha Sarakham, Thailand.*
[4]*Department, of Digital Business, School of Information and Communication Technology, University of Phayao, Phayao, Thailand.*

[4]*Corresponding Author : pratya.nu@up.ac.th*

Received: 15 November 2024 Revised: 21 March 2025 Accepted: 28 March 2025 Published: 26 April 2025

***Abstract** - The prevalence of morbidity and disease among the elderly population in Thailand constitutes a significant concern that necessitates focused attention, with implications for the national budget. Hence, the objective of this research is to enhance the classification model assessing the risk of osteoarthritis among the elderly demographic in Thailand, to evaluate the model's performance utilizing artificial intelligence technology, and to summarize and identify the risk factors associated with osteoarthritis within this population. The research is structured to compare two data sources: the northern and central regions of Thailand, comprising 354 samples from eight villages in Sop Prap Subdistrict, Sop Prap District, Lampang Province, and 368 samples from eleven villages in Chaichumphon Subdistrict, Laplae District, Uttaradit Province. To bolster the prediction accuracy of osteoarthritis risk among the elderly, the researchers have developed an advanced methodology comprising two components: a combined learning model that integrates XGBoost, LightGBM, Stacking, Bagging, and Voting techniques and an embedded learning model that includes Random Forest, Support Vector Machine, Logistic Regression, and Decision Tree methodologies. The findings indicate that the model constructed utilizing XGBoost and LightGBM techniques achieves an accuracy of up to 99.15%. Consequently, this model may be adapted for integration into a mobile application and developed as a strategy for the care of the elderly population in Thailand.*

***Keywords** - AI for healthcare, Applied informatics for medical, Medical informatics, Medical innovations, Screening osteoarthritis.*

1. Introduction

Thailand is transitioning into a comprehensive aging society, which significantly influences the health issues faced by the elderly population, particularly osteoarthritis [1], [2], [3], [4]. Overview of Osteoarthritis Issues in Thailand-Osteoarthritis represents a prevalent health concern among the elderly population in Thailand. This condition arises from cartilage degradation within the knee joint, which leads to the bones contacting one another, resulting in pain, swelling, and impaired knee mobility. Consequently, this condition adversely affects the affected individuals' quality of life and mobility. Numerous studies have emphasized the significance of osteoarthritis. Notable examples include: "The Epidemiology of Osteoarthritis of the Knee in Elderly Patients in Thailand": This study presents an epidemiological analysis of 392 elderly Thai patients with knee OA, comprising 86 males and 306 females, with an average age of 67.8 years. The prevalence of knee OA was found to range from 34.5% to 45.6% [1]. "Using a Discrete Choice Experiment to Elicit Patients' Preferences and Willingness to Pay for Knee Osteoarthritis Treatments in Thailand": This research examined the preferences of Thai patients and their willingness to pay for treatments of knee osteoarthritis. The study emphasized the significance of patient-centered care in managing knee osteoarthritis in Thailand [2]. "Prevalence, Patterns, and Risk Factors of Knee Osteoarthritis in Thai Monks": This study examined the prevalence, patterns, and risk factors of knee osteoarthritis among Thai monks, offering insights into how lifestyle and occupational activities affect knee health [3]. "Causal Model of Health Literacy in Thai

Older Adults with Knee Osteoarthritis": This research investigated the relationships among provider-patient communication, patient engagement, cognitive function, knowledge of knee OA, internet usage, social support, social participation, and health literacy in Thai older adults with knee OA [4].

Context and Definition of Osteoarthritis-Osteoarthritis (primary OA knee) is a common health problem in the elderly caused by long-term use of the knee joints [5], [6]. However, osteoarthritis is more common in young people, especially in middle age, and tends to increase continuously. The cause may not be age but overweight and overusing or misusing the knee joint, which causes the problem of osteoarthritis earlier than usual, with pain and stiffness in the knee joint, up to severe pain. Proper self-care can reduce the risk and issues that may lead to osteoarthritis. Knee Osteoarthritis is a disease caused by the deterioration of the knee cartilage, including the shape, structure, and function of the joint cartilage and the bones near the joint, and wear and tear with age. When no cartilage is covered, the bones collide while bearing weight, causing knee pain, swelling, and joint stiffness, becoming more severe over time. The knee will become deformed, and it will be impossible to perform daily activities typically [7].

The causes of osteoarthritis can be divided into two leading causes according to the nature of the occurrence: 1) Causes from primary degeneration (primary knee osteoarthritis), a condition caused by the deterioration of the cartilage surface due to age. Examples include increasing age, which is the most common cause. As age, the cartilage, muscles, and nerves deteriorate. Osteoarthritis begins at age 40. At age 55 and older, knee pain and osteoarthritis begin. Up to 40 percent of people will develop osteoarthritis at age 60. Gender: Found in females 2-3 times more than males due to lack of estrogen hormone, which prevents the degeneration of knee cartilage. Genetics: Studies have found that genetics is also involved. Patients whose family members or relatives have osteoarthritis have a higher chance of developing osteoarthritis.

Overweight: People who are overweight or obese have a higher chance of developing osteoarthritis because an increase in weight of 0.5 kilograms increases the force applied to the knee joint by 1-5 kilograms. In addition, excessive fat cells affect cartilage and bone cells, causing osteoarthritis to occur faster. The second cause is secondary knee osteoarthritis or degeneration with known causes. These include two leading causes: traumatic impact and certain diseases, such as rheumatoid arthritis, gout, and septic arthritis. These diseases affect organs outside the knee joint, including various types of arthritis (inflammatory joint disease). AI technology in health applications is therefore significant for the care of older people [8], [9], [10]. Here are some interesting research studies focusing on osteoarthritis (OA) in older adults in Thailand, highlighting the importance and new advances in the disease: "Health Literacy and Self-management among Older Persons with Knee Osteoarthritis Stage 3" [11]-This study presents descriptive research aimed at examining health literacy, self-management, and the relationship between health literacy and self-management in elderly individuals with stage 3 knee osteoarthritis. Purposive sampling was employed to select elderly participants diagnosed with stage 3 knee osteoarthritis. Data collection involved a personal data questionnaire, a health literacy questionnaire designed for elderly individuals with knee osteoarthritis, and a self-management behavior questionnaire tailored to this population. The analysis utilized descriptive statistics and Pearson's product-moment correlation coefficient. The study's results indicated that the mean health literacy score among elderly individuals with stage 3 knee osteoarthritis was at a moderate level.

Additionally, the mean self-management score was also at a moderate level. A statistically significant positive relationship was found between health literacy and self-management at a moderate level. These findings provide essential information for nurses and health professionals to enhance health literacy among elderly individuals with knee osteoarthritis, thereby improving self-management [11].

“A comprehensive health education plus monitoring support program for older adults with knee osteoarthritis coexisting with overweight and type 2 diabetes” [12]-This research is highly intriguing as it constitutes a study aimed at examining the patient outcomes associated with a comprehensive health education program supplemented by village health volunteer monitoring support, specifically targeting older adults suffering from knee osteoarthritis, who are also overweight and diagnosed with type 2 diabetes.

They aim to study a variety of diseases that may affect osteoarthritis. Additionally, research has been conducted on the impact of osteoarthritis in rural areas of Thailand. The study titled "Effectiveness of the Health Literacy Promoting Program Among Older Adults with Knee Osteoarthritis in the Rural Phayao Province” [13] underscores the significance of the osteoarthritis issue in Thailand. The above significance drives researchers to apply AI technology in health applications with three main objectives. The first objective was to improve the classification model of knee osteoarthritis risk in the elderly population of Thailand. The second objective is to evaluate the model's performance using AI technology. The third objective was to summarize and identify risk factors for osteoarthritis in the elderly population of Thailand.

2. Materials & Methods

2.1. Research Scope and Data Collection

The target was to expand the scope of the research from one area to two areas, where the original data used the population and sample of the northern region of Thailand, consisting of 354 samples from eight villages in the Sop Prap

Subdistrict, Sop Prap District, Lampang Province, Thailand. The researchers collected additional data in the central area of Thailand, consisting of 368 samples from eleven villages in Chaichumphon Subdistrict, Laplae District, Uttaradit Province, Thailand. The two target groups are summarized in Tables 1 and 2. In addition, the sample was referenced using the Krejcie and Morgan method, and each village was sampled using the proportional method.

Table 1. Sample Group of 1st Data Source

Village No.	Population	Gender		Samples
		Female	Male	
No. 4 Pae	592 (13.19%)	301 (6.71%)	291 (6.49%)	47 (1.04%)
No. 5 Thung	797 (17.76%)	407 (9.07%)	390 (8.69%)	63 (1.40%)
No. 6 Watthana	445 (9.92%)	218 (4.86%)	227 (5.06%)	35 (0.78%)
No. 9 Thung Phatthana	926 (20.64%)	454 (10.12%)	472 (10.52%)	73 (1.63%)
No. 10 Mai Watthana	265 (5.91%)	135 (3.01%)	130 (2.90%)	21 (0.47%)
No. 11 Thung Charoen	546 (12.17%)	272 (6.06%)	274 (6.11%)	43 (0.96%)
No. 12 Thung Ruang Thong	587 (13.08%)	299 (6.66%)	288 (6.42%)	46 (1.03%)
No. 15 Hong Pu Samakkhi	329 (7.33%)	170 (3.79%)	159 (3.54%)	26 (0.58%)
Total:	4,487 (100%)	2,256 (50.28%)	2,231 (49.72%)	354 (7.89%)

Table 2. Sample Group of 2nd Data Source

Village No.	Population	Gender		Samples
		Female	Male	
No. 1 Nam Sai	729 (8.32%)	352 (4.02%)	377 (4.31%)	31 (0.35%)
No. 2 Ton Kham	774 (8.84%)	390 (4.45%)	384 (4.39%)	33 (0.37%)
No. 3 Hong Sung	1,282 (14.64%)	673 (7.69%)	609 (6.95%)	54 (0.62%)
No. 4 Khum	920 (10.51%)	510 (5.82%)	410 (4.68%)	39 (0.44%)
No. 5 Na Thale	1,800 (20.55%)	949 (10.84%)	851 (9.72%)	76 (0.86%)
No. 6 Cham Pa Wai	855 (9.76%)	437 (4.99%)	418 (4.77%)	36 (0.41%)
No. 7 Huai Chang	339 (3.87%)	155 (1.77%)	184 (2.10%)	14 (0.16%)
No. 8 Rong Yang	632 (7.22%)	319 (3.64%)	313 (3.57%)	27 (0.30%)
No. 9 Nam Sai Tai	546 (6.24%)	282 (3.22%)	264 (3.01%)	23 (0.26%)
No. 10 Pa Sak	429 (4.90%)	236 (2.69%)	193 (2.20%)	18 (0.21%)
No. 11 Pak Thang	451 (5.15%)	251 (2.87%)	200 (2.28%)	19 (0.22%)
Total:	8,757 (100%)	4,554 (52.00%)	4,203 (48.00%)	368 (4.20%)

Table 1 presents the population and sample from the first source, which has a total population of 4,487, divided into 2,256 females (50.28%) and 2,231 males (49.72%). According to the Krejcie and Morgan principle, 354 people must be sampled, or 7.89 percent of the population.

Table 2 demonstrates the population and sample from the second source, which has a total population of 8,757, divided into 4,554 females (52.00%) and 4,203 males (48.00%). According to the Krejcie and Morgan principle, 368 people must be sampled, or 4.20 percent of the population.

2.2. Research Tools

The main research instrument used for data collection was a questionnaire divided into three parts: Part 1, general information; Part 2, risk assessment of the Oxford Knee Score (OKS) using the Thai version [14], [15]; and Part 3, food consumption behavior. Seven experts evaluated this questionnaire, including five public health experts and two academics from academic institutions. Details of the three parts are shown in Tables 3 to 6.

Table 3. General Information Query

Questions and Options
Consent to provide your data for this research. ❒ Consent, ❒ Not consent
Gender ❒ Male ❒ Female
Year of Birth ....
Weight .... Kilograms Height .... Centimeters
Occupation ❒ Agriculturist ❒ Enterprise employee ❒ General employee ❒ Government official ❒ Housekeeper ❒ Merchant/Own business ❒ Student ❒ Unemployed
Education level ❒ Primary level ❒ Secondary level ❒ Diploma level ❒ Bachelor level

❒ Higher than a bachelor's degree
Number of family members ....
Congenital disease (as diagnosed by a doctor) ❒ No congenital disease ❒ Blood pressure disease ❒ Cardiovascular disease ❒ Diabetes ❒ Gout ❒ Hyperlipidemia disease ❒ Kidney disease ❒ Osteoarthritis ❒ Rheumatoid disease

Table 4. The Oxford Knee Score (OKS) [16]

Questions and Options
During the past 4 weeks........ 1. How would you describe the pain you <u>usually</u> have in your knee? ❒ None (4) ❒ Very mild (3) ❒ Mild (2) ❒ Moderate (1) ❒ Severe (0)
2. Have you had any trouble with washing and drying yourself (all over) <u>because of your knee</u>? ❒ No trouble at all (4) ❒ Very little trouble (3) ❒ Moderate trouble (2) ❒ Extreme difficulty (1) ❒ Impossible to do (0)
3. Have you had any trouble getting in and out of a car or using public transport <u>because of your knee</u>? (whichever you would tend to use) ❒ No trouble at all (4) ❒ Very little trouble (3) ❒ Moderate trouble (2) ❒ Extreme difficulty (1) ❒ Impossible to do (0)
4. For how long have you been able to walk before the <u>pain from your knee</u> becomes severe? (with or without a stick) ❒ No pain/More than 30 minutes (4) ❒ 16 to 30 minutes (3) ❒ 5 to 15 minutes (2) ❒ Around the house <u>only</u> (1) ❒ Not at all - pain severe when walking (0)
5. After a meal (sitting at a table), how painful has it been for you to stand up from a chair <u>because of your knee</u>? ❒ Not at all painful (4) ❒ Slightly painful (3) ❒ Moderately painful (2) ❒ Very painful (1) ❒ Unbearable (0)
6. Have you been limping when walking <u>because of your knee</u>? ❒ Rarely/never (4) ❒ Sometimes, or just at first (3) ❒ Often, not just at first (2) ❒ Most of the time (1) ❒ All of the time (0)
7. **Could** you kneel down and get up again afterwards? ❒ Yes, easily (4) ❒ With little difficulty (3) ❒ With moderate difficulty (2) ❒ With extreme difficulty (1) ❒ No, Impossible (0)
8. Have you been troubled by <u>pain from your knee</u> in bed at night? ❒ No nights (4) ❒ Only 1 or 2 nights (3) ❒ Some nights (2) ❒ Most nights (1) ❒ Every night (0)
9. How much has <u>pain from your knee</u> interfered with your usual work (including housework)? ❒ Not at all (4) ❒ A little bit (3) ❒ Moderately (2) ❒ Greatly (1) ❒ Totally (0)
10. Have you felt that your knee might suddenly 'give way' or let you down? ❒ Rarely/never (4) ❒ Sometimes, or just at first (3) ❒ Often, not just at first (2) ❒ Most of the time (1) ❒ All of the time (0)
11. **Could** you do the household shopping <u>on your own</u>? ❒ Yes, easily (4) ❒ With little difficulty (3) ❒ With moderate difficulty (2) ❒ With extreme difficulty (1) ❒ No, Impossible (0)
12. **Could** you walk down one flight of stairs? ❒ Yes, easily (4) ❒ With little difficulty (3) ❒ With moderate difficulty (2) ❒ With extreme difficulty (1) ❒ No, Impossible (0)

Table 4 shows the English version of the Oxford Knee Score (OKS). Each question has five options, and the value of each option is displayed. After respondents provided their information, researchers summed and calculated their risk of osteoarthritis using criteria as shown in Table 5.

Table 6 shows questions about food consumption behavior consisting of 13 questions with five levels (scores) of information criteria as follows: Level 1 contains a value of 0 points, which means it has never been consumed. Level 2 includes a value of 1 point, which means it has been consumed 1-2 times per week. Level 3 contains a value of 2 points, meaning it has been consumed 3-4 times weekly.

Table 5. The Oxford Knee Score (OKS) Criteria

Score Range	Score Grading	Severity Level
0-19	May indicate severe knee osteoarthritis.	Level 4
20-29	May indicate moderate to severe knee osteoarthritis.	Level 3
30-39	May indicate mild to moderate knee osteoarthritis.	Level 2
40-48	May indicate satisfactory joint function.	Level 1

Table 6. Food consumption behavior questionnaire

Stage	Questions for food consumption behavior
Q1.	Consistence in consuming high-fiber foods such as eggplant, passion fruit, guava, etc.
Q2.	Consistency in consuming low-fiber foods such as cucumber, lychee, longan, etc.
Q3.	Consistency in fast food consumption.
Q4.	Consistency in consuming fried foods.
Q5.	Consistency in consuming 6-8 glasses of water per day.
Q6.	Consistency in consuming soft drinks and syrups.
Q7.	Consistence in consuming tea and coffee.
Q8.	Consistency in consuming alcohol.
Q9.	Consistency in consuming milk.
Q10.	Consistency in consuming instant noodles and canned foods.
Q11.	Consistency in consuming snacks.
Q12.	Consistency in consuming fermented foods.
Q13.	Consistency in consuming spicy foods.

Level 4 comprises a value of 3 points, which means it has been consumed 5-6 times per week. Level 5 retains a value of 4 points, meaning it has been consumed regularly daily.

2.3. Modelling

To develop a reasonable model for improving the prototype of predicting the risk of osteoarthritis, the researchers used an algorithm to find the best parameters for the effectiveness and the best model, as shown in Table 7, by integrating seven prediction improvement techniques into two sections.

The 1st section is the ensemble learning models, including XGBoost [16], LightGBM [17], Stacking, Bagging, and Voting techniques [18].

The parameter settings are shown in Table 8. The 2nd section is embedded learning, including Random Forest, Support Vector Machine, Logistic Regression, and Decision Tree.

Table 7. Algorithm

```
BEGIN
  Load dataset
        Define X is the members of rows 1, 2, 3,...,m without the class label
        Define y is the class label and members of row1, 2, 3,...,m
  Preprocess data to clean and remove the data that useless data
  Define the base classifiers is the members of Bcn=1, 2, 3,...,n
  Define the ensemble classifiers is the members of Ecn=1, 2, 3,...,n
  Define the parameter grids is the PARAM_GRIDS for Bagging, XGBoost, and LightGBM
  BEST_MODELS = { }
     FOR (Ecn, model) IN MODELS DO
       IF Ecn IN PARAM_GRIDS THEN
         grid_search = CREATE GridSearch(model, PARAM_GRIDS[Ecn])
         FIT grid_search ON (X, y)
         BEST_MODELS[Ecn] = grid_search BEST PARAMETERS
         CALCULATE SCORE
         y_predict = PREDICT using grid_search
         CALCULATE the classification_report for (y, y_ predict) and accuracy
       ELSE
         FIT model ON (X, y)
         y_ predict = PREDICT using model
         CALCULATE the classification_report for (y, y_ predict) and accuracy
   PRINT BEST_MODELS
END
```

Table 8. The parameter settings

Classifiers	Best Parameters	Best Score
Bagging	estimator__max_depth = 15, n_estimators = 100	0.7747
XGBoost	learning_rate = 0.2, max_depth = 3, n_estimators = 200	0.8187
LightGBM	learning_rate = 0.1, max_depth = -1, n_estimators = 100	0.8180

2.4. Model Performance and Evaluation Tools

The tools and techniques to evaluate the model performance include 5-fold Cross-Validation, the Confusion Matrix, and four indicators: accuracy, precision, recall, and f1-score values. The 5-fold Cross-Validation technique splits the test data into five equal parts.

Four parts of the data are extracted for each round to build the model, and the remaining part is used for testing. The process is repeated for five rounds, and the results are averaged. Each assessment round uses a confusion matrix tool, which compares the results with the facts derived from the collected data.

Four metrics support the confusion matrix and describe the model performance. The accuracy value represents the model's capability, the precision value represents its predictability on a class-by-class basis, and the recall value represents its accuracy on a class-by-class basis.

Finally, the f1-score is a harmonic means of precision and recall created to serve as a single metric to measure the performance of a model. The results of the tests in each technique are shown in Table 14.

3. Results

3.1. Context of the Population

The summary of the sample context mainly presents age and occupation, which are summarized in Tables 9 and 10.

Table 9. The collected samples were classified according to age and data sources

Age	Data Sources		Total
	1st Data Source	2nd Data Source	
Age less than 30 years	0 (0.00%)	58 (8.03%)	58 (8.03%)
Age 30 – 40 years	13 (1.80%)	56 (7.76%)	69 (9.56%)
Age 41 – 50 years	35 (4.85%)	43 (5.96%)	78 (10.80%)
Age 51 – 60 years	97 (13.43%)	69 (9.56%)	166 (22.99%)
Age 61 – 70 years	158 (21.88%)	70 (9.70%)	228 (31.58%)
Age 71 – 80 years	40 (5.54%)	39 (5.40%)	79 (10.94%)
Age 81 – 90 years	9 (1.25%)	28 (3.88%)	37 (5.12%)
Age over 90 years	2 (0.28%)	5 (0.69%)	7 (0.97%)
Total:	354 (49.03%)	368 (50.97%)	722 (100%)

Table 10. The collected samples were classified according to occupation and data sources

Occupation	Data Sources		Total
	1st Data Source	2nd Data Source	
Agriculturist	203 (28.12%)	179 (24.79%)	382 (52.91%)
Enterprise employee	31 (4.29%)	39 (5.40%)	70 (9.70%)
General employee	67 (9.28%)	57 (7.89%)	124 (17.17%)
Government official	2 (0.28%)	8 (1.11%)	10 (1.39%)
Housekeeper	4 (0.55%)	14 (1.94%)	18 (2.49%)
Merchant/Own business	38 (5.26%)	48 (6.65%)	86 (11.91%)
Unemployed	9 (1.25%)	23 (3.19%)	32 (4.43%)
Total:	354 (49.03%)	368 (50.97%)	722 (100%)

Table 11. Osteoarthritis risk assessment

Stage of Osteoarthritis	Data Sources		Total
	1st Data Source	2nd Data Source	
Level 4: Severe	19 (2.63%)	200 (27.70%)	219 (30.33%)
Level 3: Moderate	143 (19.81%)	30 (4.16%)	173 (23.96%)
Level 2: Mild	19 (2.63%)	125 (17.31%)	144 (19.94%)
Level 1: Normal	173 (23.96%)	13 (1.80%)	186 (25.76%)
Total:	354 (49.03%)	368 (50.97%)	722 (100%)

Table 9 displays the sample data collection classified by age and data source. Most samples were aged between 61 and 70, with 228 samples, or 31.58%, followed by those aged between 51 and 60, with 166 samples, or 22.99%. However, the least number of the sample was older than 90, with seven samples, or 0.97%. Table 10 demonstrates the collection of sample data classified by occupation and source of data.

Most informants were agriculturists, with 382 samples, or 52.91 percent, followed by those who worked as general employees, with 124 samples, or 17.17 percent. However, the fewest samples were government officials, with ten samples, or 1.39 percent.

3.2. Results of Knee Osteoarthritis Risk Assessment

Table 11 depicts the results of the analysis of the severity of osteoarthritis. It was found that the sample groups in each area had different levels of disease severity. Therefore, it is advantageous to analyse and compare.

Overall, most of the data obtained showed a large number of people at risk of osteoarthritis (severe), 219 cases, or 30.33 percent, while the sample group with mild illnesses had the fewest, 144 cases, or 19.94 percent.

3.3. Consumption Behaviour Analysis Results

Table 12. The results of consumption behavior analysis from 1st data source

Stage	Count					Average	S.D.	Interpretation
	Score 4	Score 3	Score 2	Score 1	Score 0			
Q1.	7	6	86	197	58	1.17	0.79	Consumed 1-2 times/week
Q2.	3	1	134	163	53	1.26	0.74	Consumed 1-2 times/week
Q3.	1	15	54	105	179	0.74	0.89	Never been consumed
Q4.	1	25	87	130	111	1.08	0.93	Consumed 1-2 times/week
Q5.	350	3	1	0	0	3.99	0.14	Consumed 5-6 times/week
Q6.	2	1	22	130	199	0.52	0.68	Never been consumed
Q7.	35	8	27	77	207	0.83	1.27	Never been consumed
Q8.	3	3	20	120	208	0.51	0.72	Never been consumed
Q9.	150	26	43	90	45	2.41	1.54	Consumed 3-4 times/week
Q10.	0	1	11	133	209	0.45	0.57	Never been consumed
Q11.	2	0	44	156	152	0.71	0.72	Never been consumed
Q12.	0	4	48	150	152	0.73	0.73	Never been consumed
Q13.	12	58	116	124	44	1.63	1.01	Consumed 1-2 times/week

Table 13. The results of consumption behavior analysis from 2nd data source

Stage	Count					Average	S.D.	Interpretation
	Score 4	Score 3	Score 2	Score 1	Score 0			
Q1.	34	65	167	87	15	2.04	0.97	Consumed 3-4 times/week
Q2.	37	36	149	138	8	1.88	0.97	Consumed 1-2 times/week
Q3.	12	26	38	186	106	1.05	0.98	Consumed 1-2 times/week
Q4.	2	63	106	191	6	1.63	0.80	Consumed 1-2 times/week
Q5.	356	11	1	0	0	3.96	0.20	Consumed 5-6 times/week
Q6.	27	18	63	201	59	1.33	1.04	Consumed 1-2 times/week
Q7.	134	15	61	82	76	2.13	1.59	Consumed 3-4 times/week
Q8.	45	45	49	82	147	1.35	1.42	Consumed 1-2 times/week
Q9.	79	60	108	120	1	2.26	1.14	Consumed 3-4 times/week
Q10.	41	52	42	165	68	1.55	1.25	Consumed 1-2 times/week
Q11.	42	19	70	136	101	1.36	1.25	Consumed 1-2 times/week
Q12.	11	32	38	234	53	1.22	0.90	Consumed 1-2 times/week
Q13.	50	109	93	72	44	2.13	1.22	Consumed 3-4 times/week

Table 12 shows the data of all 354 sample groups, which can be analysed to show that the sample groups took good care of themselves; for example, Question 5 asked about drinking eight glasses of water per day and not eating snacks in Question 11. Table 13 shows the data of 368 samples from the second data source, which can be analysed to show that the sample group took care of themselves appropriately. The data from Tables 12 and 13 were used to develop a mathematical model to create a classification risk prediction model that may be induced by consumer behaviour. The prediction results will have four answers: the four severity levels of osteoarthritis obtained from the assessment results.

3.4. Model Constructed and Model Performance

Table 14. Model Constructed and Model Performance

Class	Classifiers / Indicators					
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
	XGBoost			LightGBM		
Level 1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Level 2	1.0000	0.9474	0.9730	1.0000	0.9474	0.9730
Level 3	0.9795	1.0000	0.9896	0.9795	1.0000	0.9896
Level 4	1.0000	0.8889	0.9412	1.0000	0.8889	0.9412
Macro Avg.	0.9949	0.9591	0.9759	0.9949	0.9591	0.9759
Weighted Avg.	0.9917	0.9915	0.9913	0.9917	0.9915	0.9913
Accuracy	0.9915			0.9915		

	Stacking			Bagging		
Level 1	0.9611	1.0000	0.9802	0.9609	0.9942	0.9773
Level 2	0.0000	0.0000	0.0000	1.0000	0.8947	0.9444
Level 3	0.8266	1.0000	0.9051	0.9660	0.9930	0.9793
Level 4	0.0000	0.0000	0.0000	1.0000	0.5556	0.7143
Macro Avg.	0.4469	0.5000	0.4713	0.9817	0.8594	0.9038
Weighted Avg.	0.8059	0.8952	0.8470	0.9671	0.9660	0.9629
Accuracy	0.8952			0.9660		
	Voting			Random Forest		
Level 1	0.9005	0.9942	0.9451	0.8205	0.9412	0.8767
Level 2	1.0000	0.3684	0.5385	0.0000	0.0000	0.0000
Level 3	0.931	0.9441	0.9375	0.8710	0.9310	0.9000
Level 4	1.0000	0.5556	0.7143	0.0000	0.0000	0.0000
Macro Avg.	0.9579	0.7156	0.7838	0.4229	0.4681	0.4442
Weighted Avg.	0.9233	0.9178	0.9083	0.7594	0.8429	0.7987
Accuracy	0.9178			0.8301		
	Support Vector Machine			Logistic Regression		
Level 1	0.6970	0.6765	0.6866	0.7586	0.6471	0.6984
Level 2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Level 3	0.6216	0.7931	0.6970	0.6410	0.8621	0.7353
Level 4	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Macro Avg.	0.3296	0.3674	0.3459	0.3499	0.3773	0.3584
Weighted Avg.	0.5961	0.6571	0.6222	0.6340	0.6714	0.6439
Accuracy	0.6940			0.6573		
	Decision Tree					
Level 1	0.7895	0.8824	0.8333			
Level 2	0.0000	0.0000	0.0000			
Level 3	0.8621	0.8621	0.8621			
Level 4	0.0000	0.0000	0.0000			
Macro Avg.	0.4129	0.4361	0.4239			
Weighted Avg.	0.7406	0.7857	0.7619			
Accuracy	0.7990					

The constructed models and model performance are reported, with the results classified by each classifier summarized in Table 14. The models with XGBoost and LightGBM techniques had the highest accuracy, with a value of 99.15%.

The second highest accuracy is constructed with the Bagging technique, with a value of 96.60%. The model built with the Logistic Regression technique has the lowest accuracy, with a value of 65.73%.

4. Discussion

From the results, it can be discussed that when using a base classification classifier with the working principle of embedded learning models, the problem of imbalanced data is found because the amount of data in each class is very different, as displayed in Table 11.

Therefore, the method of finding the best parameter value of the classifier is found, namely Bagging, XGBoost, LightGBM, and a classifier that works well with the Imbalance data problem.

5. Conclusion

Prevention and surveillance are the best forms of health care. Furthermore, older adults are also considered a vulnerable group that requires special care. For this reason, it is necessary to manage processes or modern technologies to sustain and support the care of older people. Therefore, this research aims to improve the classification model of knee osteoarthritis risk in the elderly population of Thailand, evaluate the model's performance using AI technology, and summarize and identify risk factors for osteoarthritis in the elderly population of Thailand. The research was designed to compare two data sources: Thailand's northern and central regions. Research data has been upgraded, and the scope of research data has been expanded further. The samples consist of 354 samples from eight villages in the Sop Prap Subdistrict, Sop Prap District, Lampang Province, Thailand, and 368 samples from eleven villages in Chaichumphon Subdistrict, Laplae District, Uttaradit Province, Thailand. To improve the prediction of the risk of osteoarthritis in older adults, researchers have developed a more advanced technique that consists of two sections: the ensemble learning models, including XGBoost, LightGBM, Stacking, Bagging, and Voting techniques, and the embedded learning, including Random Forest, Support Vector Machine, Logistic Regression, and Decision Tree. The researchers found several similarities and differences. The similarities between the two areas are the context of the majority of the population in the area, which is the elderly affected by osteoarthritis. Furthermore, the occupations of the majority of the sample group are in the agricultural sector, which has a low average income. This makes it difficult for them to take care of their own health. The difference is that the two areas have populations with completely different disease levels, as shown in Table 11. The researchers found that the sample group in the North had a less severe level of illness (Osteoarthritis level 3: moderate, with 143 cases (19.81%), and Osteoarthritis level 1: normal, with 173 cases (23.96%)), while the Central region had the highest level of severe disease (Osteoarthritis level 4: severe, with 200 cases (27.70%), and Osteoarthritis level 2: mild, with 125 cases (17.31%)). There is a need to study further what happens in these two very different communities. Meanwhile, researchers developed a model to predict the risk of osteoarthritis and found that the model generated by XGBoost and LightGBM techniques had the highest accuracy, with an accuracy of 99.15%.

It is reasonable to adapt this model into a mobile application and develop it as a strategy for caring for older adults in Thailand. Therefore, the researchers are confident that this research will be of sustainable value and benefit the Thai people.

Acknowledgments

This research project was supported by the Thailand Science Research and Innovation Fund and the University of Phayao. It also received support from many advisors, academics, researchers, students, and staff. The authors thank everyone for their support and cooperation in completing this research.

References

[1] Vilai Kuptniratsaikul et al., "The Epidemiology of Osteoarthritis of the Knee in Elderly Patients Living an Urban Area of Bangkok," *Journal of the Medical Association of Thailand*, vol. 85, no. 2, pp. 154-161, 2002. [Google Scholar]

[2] Parnnaphat Luksameesate et al., "Using a Discrete Choice Experiment to Elicit Patients' Preferences and Willingness-to-Pay for Knee Osteoarthritis Treatments in Thailand," *Scientific Reports*, vol. 13, no. 1, pp. 1-9, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[3] Boonsin Tangtrakulwanich, Alan F. Geater, and Virasakdi Chongsuvivatwong, "Prevalence, Patterns, and Risk Factors of Knee Osteoarthritis in Thai monks," *Journal of Orthopaedic Science*, vol. 11, no. 5, pp. 439-445, 2006. [CrossRef] [Google Scholar] [Publisher Link]

[4] Porntip Pa-in et al., "Causal Model of Health Literacy in Thai Older Adults with Knee Osteoarthritis," *Pacific Rim International Journal of Nursing Research*, vol. 27, no. 2, pp. 303-316, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[5] E. Lawrence Hart et al., "The Relationship between Exercise and Osteoarthritis in the Elderly," *Clinical Journal of Sport Medicine*, vol. 18, no. 6, pp. 508-521, 2008. [CrossRef] [Google Scholar] [Publisher Link]

[6] Xueshan Sun et al., "Osteoarthritis in the Middle-Aged and Elderly in China: Prevalence and Influencing Factors," *International Journal of Environmental Research and Public Health*, vol. 16, no. 23, pp. 1-18, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[7] Inoshi Atukorala, and David J. Hunter, "A Review of Quality-of-Life in Elderly Osteoarthritis," *Expert Review of Pharmacoeconomics & Outcomes Research*, vol. 23, no. 4, pp. 365-381, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[8] DonHee Lee, and Seong No Yoon, "Application of Artificial Intelligence-Based Technologies in the Healthcare Industry: Opportunities and Challenges," *International Journal of Environmental Research and Public Health*, vol. 18, no. 1, pp. 1-18, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[9] Mohammed Yousef Shaheen, "Applications of Artificial Intelligence in Healthcare: A Review," *ScienceOpen Preprints*, pp. 1-8, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[10] Fei Wang, and Anita Preininger, "AI in Health: State of the Art, Challenges, and Future Directions," *Yearbook of Medical Informatics*, vol. 28, no. 1, pp. 16-26, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[11] Kanyana Meesarapee, Nattaya Suwankrauhasn, and Rojanee Chintanawat, "Health Literacy and Self-management among Older Persons with Knee Osteoarthritis Stage 3," *Nursing Research and Innovation Journal*, vol. 28, no. 3, pp. 353-369, 2022. [Google Scholar] [Publisher Link]

[12] Suparb Aree-Ue et al., "A Comprehensive Health Education Plus Monitoring Support Program for Older Adults with Knee Osteoarthritis Coexisting with Overweight and Type 2 Diabetes," *International Journal of Nursing Sciences*, vol. 9, no. 4, pp. 512-520, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[13] Porntip Pa-in, and Dao Weiangkham, "Effectiveness of the Health Literacy Promoting Program Among Older Adults with Knee Osteoarthritis in the Rural Phayao Province," *Thai Journal of Nursing*, vol. 73, no. 4, pp. 41-50, 2024. [Google Scholar] [Publisher Link]

[14] Keerati Charoencholvanich, and Boonchana Pongcharoen, "Oxford Knee Score and SF-36: Translation & Reliability for Use with Total Knee Arthroscopy Patients in Thailand," *Journal of the Medical Association of Thailand*, vol. 88, no. 9, pp. 1194-1202, 2005. [Google Scholar]

[15] Sarah L. Whitehouse et al., "The Oxford Knee Score; Problems and Pitfalls," *The Knee*, vol. 12, no. 4, pp. 287-291, 2005. [CrossRef] [Google Scholar] [Publisher Link]

[16] We Li et al., "Gene Expression Value Prediction Based on XGBoost Algorithm," *Frontiers in Genetics*, vol. 10, pp. 1-7, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[17] Yun Ju et al., "A Model Combining Convolutional Neural Network and LightGBM Algorithm for Ultra-Short-Term Wind Power Forecasting," *IEEE Access*, vol. 7, pp. 28309-28318, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[18] Pratya Nuankaew et al., "Hybrid Clustering Learning Models Based on Self-regulated Learning Model Using Unsupervised Learning by Majority Voting Techniques," *Smart Learning for a Sustainable Society Proceedings of the 7th International Conference on Smart Learning Environments*, Bangkok, Thailand, pp. 88-97, 2023. [CrossRef] [Google Scholar] [Publisher Link]