IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 19, 2021, accepted May 17, 2021, date of publication May 28, 2021, date of current version June 15, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3084623

Individual Attribute Selection Using Information Gain Based Distance for Group Classification of Elderly People With Hypertension

SUPANSA CHAISING[1], PUNNARUMOL TEMDEE[1], AND RAMJEE PRASAD[2], (Life Fellow, IEEE)

[1]Computer and Communication Engineering for Capacity Building Research Center, School of Information Technology, Mae Fah Luang University, Chiang Rai 57100, Thailand

[2]CTIF Global Capsule, Department of Business Development and Technology, Aarhus University, 7400 Herning, Denmark

Corresponding author: Punnarumol Temdee (punnarumol@mfu.ac.th)

This work was supported in part by the Thailand Research Fund through the Royal Golden Jubilee Ph.D. Program under Grant PHD/0081/2560.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the University Ethical Committee and performed in line with the Declaration of Helsinki.

ABSTRACT Attribute selection is the process of selecting relevant attributes being used in model construction to enhance model accuracy. For general medical oriented classification applications, classical attribute selection methods principally select common attributes in the dataset for all individuals. The idea of using individual attributes is proposed in this study to represent the difference among individuals for self-diagnosis. Consequently, this study proposes a new attribute selection method, called information gain based distance (IGD), for individual attribute selection, which represents an individual's health condition differently and can be used for effective classification. The proposed method combines the concept of information gain and objective distance to select individual attributes affecting classification. The IGD method is expected to provide higher classification performance than classical attribute selection methods. To assess the performance of the IGD method, classification accuracy between data with classical attribute selections and with the IGD method is compared. The case study is conducted with 971 secondary data used for group classification of elderly people with hypertension. The classification result of different classifiers was compared, including K-nearest neighbors, neural network, and naive Bayes. The comparison revealed that the classification of data with the IGD attribute selection method provided an average classification accuracy of 98.73%. In comparison, those classifications of data with classical attribute selection methods provided 62.99%, 62.99%, 62.62%, and 62.85% for information gain, Gini index, chi-squared, and decision tree, respectively. The results showed that data classification with the IGD method provided higher performance than those with the classical attribute selection methods.

INDEX TERMS Attribute selection, individual attributes, information gain, hypertension, elderly people.

I. INTRODUCTION

Generally, data classification is widely used in medical oriented applications [1]–[3]. Currently, many applications are developed for promoting the well-being of an aging society [4], such as self-monitoring [5], self-diagnosis [6], and self-management [7]. For these applications, clinical factors are normally used as mandatory factors. However, the use of clinical factors for these applications is still facing challenges because of the difficulty of gathering clinical data [8] and processing irrelevant data [9], [10]. Normally, clinical data is very diverse. It can have a small size and a low frequency of change. At the same time, it can have a large amount of irrelevant data. The existing machine learning based classifiers commonly face challenges in dealing with this diversity.

Furthermore, a principle of personalized healthcare is about providing the proper treatment for the individual patient [11]. The analysis of individual factors is thus necessary for self-diagnosis, self-monitoring, and individual treatment. For this study, the assumption is made that a different set of individual factors affect different health conditions so that they can be used for healthcare services of each

The associate editor coordinating the review of this manuscript and approving it for publication was Jad Nasreddine.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

individual differently. The main objective of this study is to propose the individual attributes selection, which can represent the difference among individuals and can be used for medical oriented classification problems effectively.

For this study, the group classification is conducted with elderly people with hypertension. Hypertension is a condition of higher than normal blood pressure when blood is pumped out of the heart into the arteries. Hypertension is a long-term medical condition that causes around 10.4 million deaths per year [12]. Particularly, it is expected to increase to 1.56 billion deaths by 2025 [13]. To reduce these huge numbers of deaths, hypertensive patients must seriously manage their hypertension levels according to healthcare professionals' suggestions.

Nevertheless, an increasing number of elderly hypertensive patients create a shortage of healthcare professionals. To aid the shortage of healthcare providers, personalized healthcare applications designed for hypertension management through digital technology have been applied for supporting elderly people with hypertension and their healthcare providers [14], [15]. According to healthcare professionals' suggestions, a lifestyle change is recommended for hypertension management [16]. Generally, healthcare providers provide healthcare recommendations to adjust patient lifestyle behaviors based on individual health conditions representing different abilities of hypertension control. Therefore, the classification of hypertension controllability among individuals is important for providing proper recommendations.

For this study, the individual attributes affecting the classification of elderly groups are selected based on individual abilities to control hypertension. More specifically, this study proposes a novel attribute selection method called information gain based distance (IGD) for individual attribute selection to obtain better classification performance. The IGD method is a modification of information gain to the original objective distance [17], [18]. The objective distance is used for determining a distance between the current health status and the expected health status to achieve the target goals for individuals. The information gain can select the important factors by measuring reductions in entropy. For this study, the combination reveals the potential attribute selection by considering the individual health conditions having high entropy, called individual attributes, resulting in better classification performance. To evaluate the performance, the comparisons of group classification with different classifiers using classical attribute selection methods and the proposed attribute selection method are conducted.

II. LITERATURE REVIEW

Attribute or feature selection is the process of selecting a subset of relevant attributes or variables from all features by eliminating the redundant and irrelevant attributes in a dataset [19]. The selected subset of relevant attributes will be used in predictive model building. The existing feature selection methods have been used to select relevant features or attributes to improve classification accuracy in diverse medical problems. For example, a metaheuristic algorithm [20], information gain [21], and chi-squared [22] were proposed to determine the relevant attributes in the heart disease prediction problem by reducing the number of irrelevant attributes obtained from patients. In the coronary artery disease (CAD) prediction problem, a hybrid Particle swarm optimization based Extreme learning machine (PSO-ELM) was proposed as a model for the diagnosis of CAD. A feature selection algorithm named Fisher was used to determine more discriminative feature subsets to enhance the performance of the proposed model [23]. In addition, a heuristic approach for feature selection based on information gain and Gini index was proposed to select the disease-related features for CAD prediction [24]. Besides, the decision tree algorithm was proposed to select the best features for Parkinson's disease predictions [25]. Medical datasets often have large attribute sets where some attributes are insignificant to some patients, so eliminating the insignificant attributes is important for improving the performance of classifiers [26]. From those studies, all categories of attribute selection method were applied to select the relevant attributes which are the same for all individuals. Conversely, the concept underlying this study is that to select the relevant attributes which are different for everyone.

Generally, attribute selection methods are divided into three categories: filter, wrapper, and embedded methods [27], [28]. Filter methods depend on the characteristics of attributes to select attribute subsets without using any machine learning algorithm. Examples of filter methods are chi-squared, information gain, Gini index, etc. Wrapper methods use a predictive machine learning algorithm to select the subsets of attributes. Examples of wrapper methods are forward feature selection, backward feature elimination, etc. Embedded methods perform attribute selection as part of the model construction combining the qualities of filter and wrapper methods, which are implemented by algorithms that have their own built-in attribute selection methods. Examples of embedded methods are a decision tree algorithm, random multinomial logit, etc. Among the three categories of attribute selection methods, this study only focuses on the filter method. The reason is that this study aims to select the relevant attributes for each individual depending on the actual effect of attributes toward classification, without using any machine learning algorithm in the attribute selection process.

Currently, the attribute selection method with instance-level information is widely introduced. For example, the saliency-based feature selection (SFS) [29] was proposed as a new feature selection method based on deep-learning saliency techniques. This method aims to rank the importance of each feature at the instance level rather than concerning the entire dataset. It helps to customize feature relevance information for each sample. Then, those features will be used by the classifier to predict a certain outcome for obtaining better classification accuracy. It can also be used in classification or regression problems, particularly in medical datasets. Similarly, this study is also interested in selecting the individual attributes based on instance-level information for obtaining better classification accuracy. However, this individual relevant attribute subset is selected without using any machine learning algorithm.

To obtain individual relevant attributes, this study proposes the IGD method as a novel attribute selection method to select individual relevant attributes from the dataset. For this study, the case study is the classification of elderly people with hypertension. The selected individual attributes represent the relevant attributes affecting the development of hypertension levels for each elderly person. The IGD method is modified from information gain and combined with the concept of objective distance [17], [18]. Information gain is known as one of the most utilized attribute selection methods to choose irrelevant attributes [30], [31]. This method generally determines the attribute importance by measuring reductions in entropy, introduced by Shannon in 1948 [32]. The low entropy provides high information gain due to entropy indicates the degree of uncertainty for the impurity of the dataset. At the same time, the concept of objective distance measures the distance at the current health status representing the path for achieving the expected level of each attribute. The combination of information gain and objective distance can enhance the competency of selecting individual relevant attributes. Those attributes are selected from prioritizing the important levels of attributes based on individual health conditions to obtain the attributes affecting the classification.

For this study, the proposed IGD method will be particularly beneficial for enhancing the classification performance of the elderly group and providing personalized healthcare recommendations. This study assumes that the proposed IGD method will enhance the classification performance more than the classical attribute selection methods. To assess the effect of individual attributes on classification, this study compares classification performance between the classification of the data set obtained from the IGD attribute selection method and classical attribute selection methods.

III. RESEARCH METHODOLOGY

The research methodology of this study comprises four main processes: data collection, attribute selection with classical methods, attribute selection with IGD method, and evaluation of classification results, as shown in Figure 1. The details of each procedure are described in the following sections.

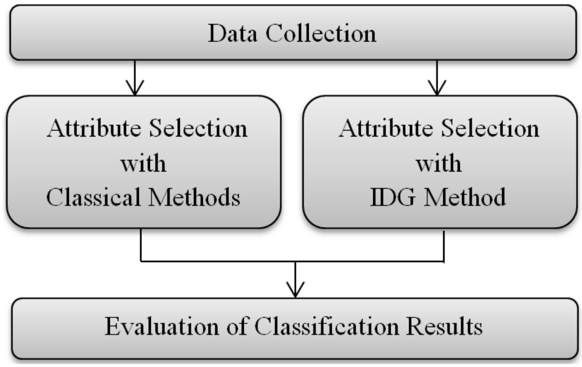

FIGURE 1. **Research methodology.**

A. DATA COLLECTION

The case study of this research is the classification of the elderly with hypertension group based on their ability to control hypertension development. The collected dataset contains 971 samples of secondary data related to clinical and lifestyle data for elderly people with hypertension and aged 65 and older. The data was collected from hospitals in Chiang Rai, Thailand. The clinical and lifestyle data gathered are total cholesterol, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, body mass index, and smoking, and indicate the risk factors associated with hypertension development as described below.

1) CLINICAL DATA

Total cholesterol (TC) refers to the total amount of cholesterol in the blood, which cannot be dissolved in blood or water. High TC levels result in the development of hypertension, particularly among elderly people. The essential components of TC include low-density lipoprotein cholesterol (LDL-C) and high-density lipoprotein cholesterol (HDL-C) [33]. LDL-C is known as bad cholesterol that accumulates on artery walls. High LDL-C levels are associated with hypertension development, so maintaining LDL-C levels below 100 mg/dL is recommended [34]. HDL-C refers to good cholesterol that eliminates bad cholesterol in the body with contributing to its removal by the liver. High HDL-C levels can prevent hypertension development and protect against heart disease [35]. Body mass index (BMI) is a measurement of a person's body fat level, which can indicate overall health. High BMI levels cause the development of hypertension and other complications, so maintaining BMI levels at a normal weight is highly suggested [36]. A normal weight for the Asian elderly group is between 23–25 kg/m^2 [37]–[39].

2) LIFESTYLE DATA

Smoking (SM) refers to inhaling and exhaling the smoke of burning plant material, which is associated with a risk factor for hypertension development. Smoking is defined as three types: a non-smoker (a person who has never smoked), an ex-smoker (a person who was once a daily smoker but currently does not smoke), and a smoker (a person who smokes any tobacco product, either daily or occasionally) [40].

The above risk factors affecting hypertension development are based on previous studies [17], [18]. However, physical activity was not considered in this study because the data gathered by hospitals related to a patient's physical activity is inadequate. Examples of secondary data related to current health data for elderly people collected from hospitals are shown in Table 1. In this table, "No" and "Yes" for the SM attribute indicate a non-smoker and a smoker, respectively. SBP indicates a level of blood pressure for each person. If a level of SBP is equal to or more than 130 mmHg, it represents a level of hypertension. If a level of SBP is less than 130 mmHg, it represents hypertension under control.

TABLE 1. Examples of secondary data related to current health data for elderly people.

Elderly no.	SBP	Current health data				
		TC	LDL-C	HDL-C	BMI	SM
1	139	235	174	44	27.34	No
2	130	228	123	74	20.51	No
3	130	226	125	77	28.62	No
4	130	214	123	63	27.82	No
5	140	179	63	48	25.33	No
6	147	202	95	34	23.73	No
7	140	201	116	64	23.94	No
8	140	141	64	56	23.13	No
9	130	120	61	40	25.88	No
10	134	163	74	65	25.21	Yes
…	…	…	…	…	…	…
971	142	205	133	39	26.35	No

The university ethical committee approved this study in accordance with the Declaration of Helsinki of 2002. Informed consent was not required in this study because the data used for an experiment was secondary. In addition, the hospitals have already approved the gathered data, so all data will never imply any individual, and there is no conflict of interest in this study.

B. ATTRIBUTE SELECTION WITH CLASSICAL METHODS

A principle of the classical attribute selection method is demonstrated in the following sections.

1) A PRINCIPLE OF THE CLASSICAL ATTRIBUTE SELECTION METHOD

A classical attribute selection method selects a relevant attribute from a set of all attributes in the whole dataset, as shown in Figure 2.

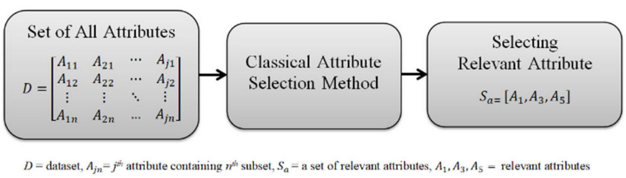

FIGURE 2. A principle of classical attribute selection method.

The principle of attribute selection underlying the classical attribute selection methods is a consideration of the whole dataset for selecting the relevant attributes. In this study, a dataset, or a set of all attributes in the whole dataset (D), as shown in Figure 2, is represented by a matrix. The $j \times n$ matrix D represents the n rows and j columns. Each element of a matrix is denoted by j^{th} attribute containing n^{th} subsets (A_{jn}). Then, the dataset is conducted by a classical attribute selection method to select relevant attributes from the whole dataset. A set of relevant attributes is denoted by S_a, for example, $S_a = [A_1, A_3, A_5]$ represents a set of relevant attributes, including A_1, A_3, A_5, which were selected by the classical method, as illustrated in Figure 2.

2) CLASSICAL ATTRIBUTE SELECTION METHODS

This study applied four different attribute selection methods to select relevant attributes affecting the classification, including information gain, Gini index, chi-squared, and decision tree.

A detail of each method is explained as follows:

a: INFORMATION GAIN

Entropy is firstly determined to obtain information gain for an attribute. The entropy is calculated as:

$$Entropy\,(D) = \sum\nolimits_{i=1}^{m} -P_i log_2 P_i \tag{1}$$

where D refers to a sample of training examples, P_i is the proportion of the examples in D that belong to the i^{th} class. Then, the information gain is calculated as:

$$Gain\,(D,j) = Entropy\,(D) - \sum\nolimits_{v\in Values(j)} \frac{|D_v|}{|D|} * Entropy\,(D_v) \tag{2}$$

where $Gain\,(D,j)$ is the information gain of an attribute j which is relative to the collection of instances D, $Values\,(j)$ represents the set of all possible classes for attribute j, and D_v is the subset of D for the attribute j, which has the value v.

b: GINI INDEX

Gini index is one of the attribute selection methods, which measures the purity of the attributes with respect to the class. In other words, the Gini index is a reduction in impurity. The minimum Gini index represents the relevant attribute. If the data set D containing samples from m classes, the Gini index, $Gini\,(D)$, is computed as:

$$Gini\,(D) = 1 - \sum\nolimits_{i=1}^{m} p_i^2 \tag{3}$$

where p_i is the relative frequency of class i in D. If a data set D is split on attribute j into v subsets, the $Gini_j\,(D)$ for attribute j after splitting is calculated as:

$$Gini_j\,(D) = \sum\nolimits_{v=1}^{n} \frac{|D_v|}{|D|} Gini(D_v) \tag{4}$$

After calculating Gini index for all subsets of an attribute, the subset that provides the minimum Gini index for that attribute is selected. The reduction in impurity that would be incurred by splitting on attribute j is calculated as:

$$Gini\,(j) = Gini\,(D) - Gini_j\,(D) \tag{5}$$

The attribute that provides the maximum reduction in purity, which has the minimum Gini index, is selected as the relevant attribute.

c: CHI-SQUARED

Chi-squared is a method of mathematical statistics, which is applied for a feature selection. This method computes the value of attributes with respect to the class attribute. The higher value represents the more relevant attributes. The chi-squared statistic is used to find if a distribution of observed frequencies is different from the expected frequencies. The formula for chi-square is calculated as:

$$X^2 = \sum \frac{(O-E)^2}{E} \tag{6}$$

where X^2 is the value of the chi-squared statistic, O is the observed frequency, E is the expected frequency.

d: DECISION TREE

Decision tree is one of the supervised machine learning algorithms, which can be used for attribute selections. This algorithm selects an attribute in each recursive step of the tree growth process and splits the sample into smaller subsets. It results in obtaining the best attributes from a given dataset for the purpose of classification. The decision tree is a tree structure including internal nodes, branches, and leaf nodes. Each internal node represents an attribute. The first internal node is indicated as the root node. Each branch represents the attribute category. Each leaf node represents a class label. Decision tree includes three algorithms. The first algorithm is ID3 (Iterative Dicotomizer 3), developed by Ross Quinlan. ID3 applies greedy algorithms to generate multiple branch trees. The trees extend to maximum size before pruning. The second algorithm is C4.5, which is an extension of ID3 by overcoming limitations of features that are required to be categorical. It defines distinct attributes for numerical features by converting the trained trees with if-then conditions. The third algorithm is C5.0, an extension of C4.5 using less space, creating smaller rule sets, and performing faster than C4.5.

C. ATTRIBUTE SELECTION WITH IGD METHOD

To select relevant attributes for better classification performance, the novel attribute selection method IGD is proposed based on applying individual attributes. Accordingly, this section demonstrates a principle of the IGD method, proposed IGD method, and sample calculation as follows.

1) A PRINCIPLE OF THE IGD METHOD

The principle of attribute selection underlying the proposed IGD method is a consideration of individual attributes for selecting the individual relevant attributes, as represented in Figure 3. The individual relevant attribute represents the actual effect of attribute toward classification for each individual. The concept of the IGD method is based on the principle of a general diagnosis normally performed by healthcare professionals that focuses on a few significant individual factors of patients affecting disease diagnosis [41]. In addition, the health data obtained from the patients is varied because each person has different health conditions, resulting in different disease diagnoses.

From Figure 3, a dataset or a set of all attributes (D) is denoted by a matrix, which is identified as the dataset used for classical attribute selection methods. The $j \times n$ matrix D represents j columns, indicating an attribute's column, and n rows, indicating attributes of each individual.

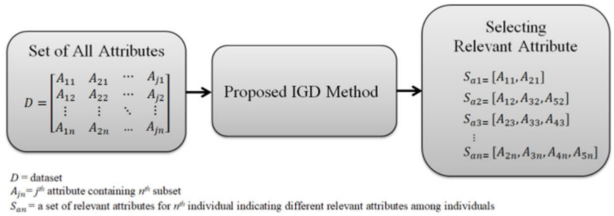

FIGURE 3. A principle of the IGD method.

Each element of a matrix is denoted by j^{th} attribute containing n^{th} subsets (A_{jn}). Then, the dataset is conducted by the proposed IGD method to select relevant attributes of individuals. A set of relevant attributes for each individual selected by the proposed method is denoted by S_{an}, which represents different relevant attributes among individuals. For example, a set of relevant attributes for the 1^{st} individual selected by the proposed method is $S_{a1} = [A_{11}, A_{21}]$, while a set of relevant attributes for the 2^{nd} individual is $S_{a2} = [A_{12}, A_{32}, A_{52}]$, as shown in Figure 3.

2) PROPOSED IGD METHOD

The attribute selection using the proposed IGD method consists of five steps, as shown in Figure 4. A detail of each step is explained as follows. A table of symbols for the proposed IGD method is presented in Appendix A.

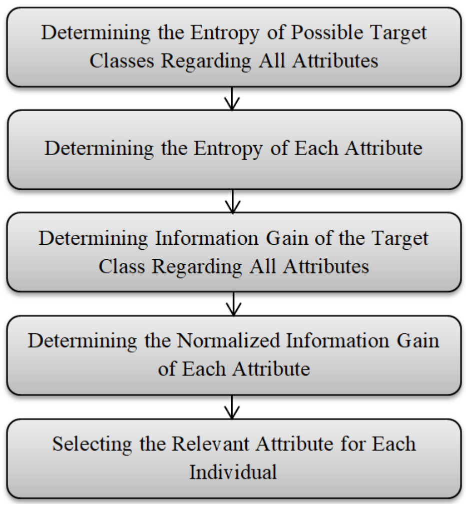

FIGURE 4. Steps of the proposed IGD method.

a: DETERMINING THE ENTROPY OF POSSIBLE TARGET CLASSES REGARDING ALL ATTRIBUTES

The equal proportion of the target class is determined. This equal proportion is assumed that the proportion of being a positive class and a negative class is equal as:

$$Ep_{+} = Ep_{-} = \frac{Nj}{Ng} \tag{7}$$

where Ep_{+} means the value of equal proportion for a positive class (+) regarding all attributes, Ep_{-} means the value of equal proportion for a negative class ($-$) regarding all attributes, Nj refers to a total number of attributes, and Ng refers to a total number of possible target classes.

Then, the target class fraction representing a chance of being a positive class or a negative class regarding all attributes is determined as:

$$fg_+ = \frac{Ep_+}{Nj} = fg_- = \frac{Ep_-}{Nj} \tag{8}$$

where fg_+ means the positive class fraction (+) and fg_- means the negative class fraction ($-$), in which both classes are balanced.

Hence, the entropy of possible target classes regarding all attributes ($E(G)$) is determined as:

$$E(G) = -fg_+ * log_2 (fg_+) - fg_- * log_2 (fg_-) \tag{9}$$

where $E(G) = 1$ indicates a high uncertainty of the class prediction representing the equivalent chance of being a positive class and a negative class.

b: DETERMINING THE ENTROPY OF EACH ATTRIBUTE

To obtain the entropy of each attribute, the objective distance [17], [18] is initially modified to compute the acceptable distance (dAB_j) and the current distance (dAC_j) of each attribute. The acceptable distance refers to the expected distance for the j attribute to achieve the expected goal. The current distance refers to the distance that must be controlled to achieve the expected distance or the expected goal for the j attribute. Specifically, conditions (i) and (ii) are predefined for the dAB_j and dAC_j calculations as follows.

Condition (i) will be considered for an attribute that has either a maximum or minimum value which is in between the acceptable range as:

Let $A_j = a_j, B_j = b_j, C_j = c_j$ and $a_j, b_j, c_j \in \mathbb{R}$.

$$B_j = \begin{cases} c_j & if\ K < k\ and\ c_j > b_j\ or\ c_j < b_j \\ b_j & otherwise \end{cases}$$

where a_j, b_j, c_j are a value of each level, K is an actual value of the target, and k is the significant value for identifying the target class.

Condition (ii) will be considered for an attribute that has minimum and maximum values which are in between the acceptable range as:

Let $A_j = a_{lj}, B_j = b_{1lj}\ or\ b_{2lj}, C_j = c_{lj}$, which $a_{lj}, b_{1lj}, b_{2lj}, c_{lj} \in R$ and $B_j \in R := [b_{1lj}, b_{2lj}]$

$$B_j = \begin{cases} b_{1lj} & if\ K \geq k\ and\ c_{lj} < b_{1lj} \\ b_{2lj} & if\ K \geq k\ and\ c_{lj} > b_{2lj} \\ c_{lj} & if\ K < k\ and\ c_{lj} < b_{1lj}, \\ & or\ K < k\ and\ c_{lj} > b_{2lj} \end{cases}$$

where $a_{lj}, b_{1lj}, b_{2lj}, c_{lj}$ are the values of each level, l is an attribute with minimum and maximum values which are in between the acceptable range, b_{1lj} and b_{2lj} are a minimum value and a maximum value for the acceptable level of j attribute, respectively. dAB_j and dAC_j are computed as:

$$dAB_j = \sqrt{(A_j - B_j)^2} \tag{10}$$

$$dAC_j = \sqrt{(A_j - C_j)^2} \tag{11}$$

where dAB_j represents the expected distance of each attribute to reach the expected goal, dAC_j represents the current distance that needs to be controlled to reach the expected distance of each attribute, A_j is the expected level of j attribute, B_j is the acceptable level of j attribute, and C_j is the current level of j attribute.

After obtaining all variables, condition (iii) is predefined for the dAC_j as:

$$dAC_j = \begin{cases} 0 & if\ C_j \geq B_j, C_j \leq B_j; B_j \notin \mathbb{R} := [b_{1lj}, b_{2lj}], \\ & or\ C_j \in B_j; B_j \in \mathbb{R} := [b_{1lj}, b_{2lj}] \\ dAC_j & otherwise \end{cases}$$

Then, the proportion of the acceptable distance (pAB_j) and the current distance (pAC_j) for each attribute are calculated as:

$$pAB_j = \frac{dAB_j}{dAB_j + dAC_j} \tag{12}$$

$$pAC_j = = \frac{dAC_j}{dAB_j + dAC_j} \tag{13}$$

Hence, the entropy of each attribute ($E(G_j)$) is computed by applying pAB_j and pAC_j, which are indicated as a positive class (pAB_{j+}) and a negative class (pAC_{j-}) respectively, as:

$$E(G_j) = -pAB_{j+} * log_2 (pAB_{j+}) - pAC_{j-} * log_2 (pAC_{j-}) \tag{14}$$

The value of $E(G_j)$ is under the predefined condition (iv) as:

$$E(G_j) = \begin{cases} 0 & if\ pAC_{j-} = 0 \\ E(G_j) & otherwise \end{cases}$$

where $E(G_j) = 0$ represents that the dAC_j has already reached the dAB_j.

c: DETERMINING INFORMATION GAIN OF THE TARGET CLASS REGARDING ALL ATTRIBUTES

The entropy of all attributes ($E(G_a)$) is used to find information gain of the target class regarding all attributes. $E(G_a)$ is computed as:

$$E(G_a) = \sum_{j=1}^{Nj} \left[E(G_j) * \left(\frac{pAB_j + pAC_j}{Nj} \right) \right] \tag{15}$$

Hence, the information gain of the target class regarding all attributes ($Gain(G, j)$) is computed as:

$$Gain(G, j) = E(G) - E(G_a) \tag{16}$$

where $Gain(G, j)$ indicates the amount of information gained from the distances of all attributes, $Gain(G, j) = 1$ if $E(G_a) = 0$, which means that all attributes have already reached the expected distance.

d: DETERMINING THE NORMALIZED INFORMATION GAIN OF EACH ATTRIBUTE

The normalized information gain of each attribute (Z_j) is computed as:

$$Z_j = \frac{E\left(G_j\right)}{Gain\left(G, j\right)} \tag{17}$$

where Z_j represents the attribute influencing the classification if $Z_j > 0$.

e: SELECTING THE RELEVANT ATTRIBUTE FOR EACH INDIVIDUAL

The relevant attribute for each individual is selected as:

$$F_j = \begin{cases} et & if\ Z_j = 0 \\ st & otherwise \end{cases} \tag{18}$$

where F_j is the j attribute, et is an eliminated attribute, st is a selected attribute.

3) SAMPLE CALCULATION

The sample calculation for selecting the individual relevant attribute using the proposed IGD method is demonstrated by selecting the relevant attributes for elderly person no.1 with hypertension, which employed information from Table 1. The individual relevant attributes represent the attributes of each elderly person that affect hypertension development. To calculate IGD, the expected level and the acceptable level of all attributes were also applied, as shown in Table 2. The expected level is the level of each attribute representing an optimal level for maintaining hypertension under control. The acceptable level is the level of each attribute that can be accepted for maintaining hypertension under control.

The expected level and acceptable level of TC, LDL-C, and HDL-C demonstrated in Table 2 were obtained from the previous work [17], [18]. The levels of BMI are based on the literature and the experts' suggestion. The acceptable BMI level should be between 23 kg/m^2 and 25 kg/m^2, which is the acceptable range for elderly people. Consequently, the minimum value of the acceptable BMI level is equal to 23 kg/m^2 if the current BMI of elderly people is less than 23 kg/m^2. The maximum value of the acceptable BMI level is equal to 25 kg/m^2 if the current BMI of elderly people is higher than 25 kg/m^2. The expected BMI level is the average for normal weight elderly people, which is equal to 24 kg/m^2. The SM category was assigned as scores: 3 (non-smoker), 2 (ex-smoker), and 1 (smoker) for the IGD calculation. The expected level and the acceptable level of SM include 3 and 2, respectively.

The five attributes used in this study are TC, LDL-C, HDL-C, BMI, and SM. The target class comprises a potential controllable individual group (PC) representing a positive class, and a potential uncontrollable individual group (PU) representing a negative class. The sample calculation of selecting the relevant attribute for elderly person no. 1 using the proposed IGD method is demonstrated below.

TABLE 2. Expected level and acceptable level of each attribute.

Attribute	Expected level	Variables (A_j)	Acceptable level	Variables (B_j)
TC	199 mg/dL	a_{TC}	219.5 mg/dL	b_{TC}
LDL-C	99 mg/dL	a_{LDL-C}	114.5 mg/dL	b_{LDL-C}
HDL-C	60 mg/dL	a_{HDL-C}	49.5 mg/dL	b_{HDL-C}
BMI	24 kg/m^2	a_{BMI}	23 kg/m^2	b_{1BMI}
			25 kg/m^2	b_{2BMI}
SM	3	a_{SM}	2	b_{SM}

a: DETERMINING THE ENTROPY OF POSSIBLE TARGET GROUPS REGARDING ALL ATTRIBUTES

Initially, the equal proportion of target groups representing an equal proportion of being PC group (Ep_+) and PU group (Ep_-) was determined by (7). The value of Ep_+ and Ep_- were equal to 2.5 as follows:

$$Ep_+ = Ep_- = \frac{5}{2} = 2.5$$

Then, the target class fraction representing a chance of being a PC group (fg_+) or a PU group (fg_-) regarding all attributes was determined by (8) as follows:

$$fg_+ = \frac{2.5}{5} \quad fg_- = \frac{2.5}{5}$$

Hence, the entropy of the target group regarding all attributes was determined by (9). The value of $E\left(G\right)$ was equal to 1 as follows:

$$E\left(G\right) = -\frac{2.5}{5} \times log_2\left(\frac{2.5}{5}\right) - \frac{2.5}{5} \times log_2\left(\frac{2.5}{5}\right) = 1$$

b: DETERMINING THE ENTROPY OF EACH ATTRIBUTE

To obtain the entropy of each attribute, TC and BMI were presented as an example of entropy calculation. Initially, the acceptable distance of TC (dAB_{TC}) and BMI (dAB_{BMI}) were computed by (10). The current distance of TC (dAC_{TC}) and BMI (dAC_{BMI}) were computed by (11). According to Table 1 and Table 2, $A_{TC} = a_{TC} = 199, A_{BMI} = a_{BMI} = 24, C_{TC} = c_{TC} = 235, C_{BMI} = c_{BMI} = 27.34$. K and k variables were considered to indicate B_j, as denoted in conditions (i) and (ii). In this study, K variable is the current SBP that indicates the current hypertension level. The current SBP of elderly person no. 1 was 140, so $K = 139$. k variable is the target level of SBP for the current SBP that needs to be maintained under this level, which $k = 130$. TC is in the condition (i) because TC has a maximum value for the acceptable level at 219.5 mg/dL, so $B_{TC} = b_{TC} = 219.5$ due to $K \geq k(139 \geq 130)$. BMI is in the condition (ii) because BMI has a minimum value (b_{1BMI}) at 23 kg/m^2 and a maximum value (b_{2BMI}) at 25 kg/m^2 of the acceptable range. The acceptable BMI level for this person was considered by $K \geq k\,(139 \geq 130)$ and $c_{BMI} > b_{2BMI}(27.34 > 25)$, so $B_{BMI} = b_{2BMI} = 25$. The dAB_{TC} and the dAC_{TC} were

TABLE 3. Entropy of each attribute for elderly person no. 1.

	Entropy value based on each attribute				
	TC	LDL-C	HDL-C	BMI	SM
Entropy	0.95	0.66	0.97	0.78	0.00

equal to 20.5 and 36 respectively as follows:

$$dAB_{TC} = \sqrt{(199-219.5)^2} = 20.5$$
$$dAC_{TC} = \sqrt{(199-235)^2} = 36$$

The dAB_{BMI} and the dAC_{BMI} are equal to 1 and 3.34 as follows:

$$dAB_{BMI} = \sqrt{(24-25)^2} = 1$$
$$dAC_{BMI} = \sqrt{(24-27.34)^2} = 3.34$$

Then, the proportion of the acceptable distance (pAB_{TC}) and the current distance (pAC_{TC}) for TC were calculated by (12) and (13), respectively. The pAB_{TC} and the pAC_{TC} were equal to 0.36 and 0.64 as follows:

$$pAB_{TC} = \frac{20.5}{20.5+36} = 0.36$$
$$pAC_{TC} = \frac{36}{20.5+36} = 0.64$$

The proportion of the acceptable distance (pAB_{BMI}) and the current distance (pAC_{BMI}) for BMI were calculated by (12) and (13), respectively. The pAB_{BMI} and the pAC_{BMI} were equal to 0.23 and 0.77 as follows:

$$pAB_{BMI} = \frac{1}{1+3.34} = 0.23$$
$$pAC_{BMI} = \frac{3.34}{1+3.34} = 0.77$$

Hence, the entropy of TC ($E(G_{TC})$) was computed by (14). The $E\,(G_{TC})$ was equal to 0.95 as follows:

$$E\,(G_{TC}) = -0.36 * log_2\,(0.36) - 0.64 * log_2\,(0.64) = 0.95$$

The entropy of BMI ($E\,(G_{BMI})$) was computed by (14). The $E\,(G_{BMI})$ was equal to 0.78 as follows:

$$E\,(G_{BMI}) = -0.23 * log_2\,(0.23) - 0.77 * log_2\,(0.77) = 0.78$$

The entropy of TC, LDL-C, HDL-C, BMI, and SM was presented in Table 3.

c: DETERMINING INFORMATION GAIN OF THE TARGET GROUP REGARDING ALL ATTRIBUTES

Initially, the entropy of all attributes ($E(G_a)$) was computed by (15). The $E\,(G_a)$ was equal to 0.67 as follows:

$$E\,(G_a) = 0.95 * \left(\frac{0.36+0.64}{5}\right) + 0.66 * \left(\frac{0.17+0.83}{5}\right) + 0.97 * \left(\frac{0.40+0.60}{5}\right) + 0.78 * \left(\frac{0.23+0.77}{5}\right) + 0 * \left(\frac{1+0}{5}\right) = 0.67$$

Hence, the information gain of the target group regarding all attributes ($Gain\,(G,j)$) was computed by (16). The $Gain\,(G,j)$ was equal to 0.33 as follows:

$$Gain\,(G,j) = 1 - 0.67 = 0.33$$

d: DETERMINING THE NORMALIZED INFORMATION GAIN OF EACH ATTRIBUTE

The normalized information gain of each attribute (Z_j) was calculated by (17). The normalized information gain of the TC attribute was presented as an example. The Z_{TC} was equal to 2.87 as follows:

$$Z_{TC} = \frac{0.95}{0.33} = 2.87$$

The normalized information gain of each attribute is shown in Table 4.

TABLE 4. The normalized information gain of each attribute for elderly person no. 1.

	Normalized information gain value based on each attribute				
	TC	LDL-C	HDL-C	BMI	SM
Normalized information gain	2.87	2.01	2.94	2.36	0.00

e: SELECTING THE RELEVANT ATTRIBUTE FOR EACH INDIVIDUAL

The relevant attributes for elderly person no. 1 were selected by (18) as follows:

$$\begin{aligned} F_{TC} &= st, && \text{as } Z_{TC} = 2.87 \\ F_{LDL-C} &= st, && \text{as } Z_{TC} = 2.01 \\ F_{HDL-C} &= st, && \text{as } Z_{HDL-C} = 2.94 \\ F_{BMI} &= st, && \text{as } Z_{BMI} = 2.36 \\ F_{SM} &= et, && \text{as } Z_{SM} = 0 \end{aligned}$$

Therefore, it can be concluded that TC, LDL-C, HDL-C, and BMI were selected as the relevant attributes for elderly person no. 1 in order to use in the elderly group classification process.

D. EVALUATION OF CLASSIFICATION RESULTS

The evaluation process for classification results is presented in Figure 5. The relevant attributes were applied in elderly group classification using different classifiers. Then, classification results obtained from each classifier were compared to evaluate the classification results. The comparison results regarding elderly group classification performance are clarified in the results and discussion section.

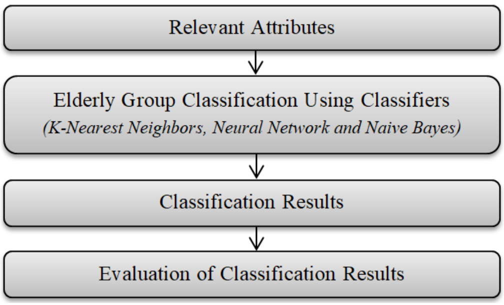

FIGURE 5. The evaluation process for classification results.

The classifiers were applied to compare results between relevant attribute selections with classical attribute selection methods and with the IGD method. The target group for classification includes two groups: potentially controllable individuals (PC) and potentially uncontrollable individuals (PU). PC represents an individual who has SBP < 130 mmHg because this level has been recommended as the optimal target blood pressure for elderly people [42]–[44]. PU represents an individual who has SBP ≥ 130 mmHg. Then, the predicted group and the actual group were compared to obtain an accuracy of classification. The classifiers employed in this study are K-nearest neighbors (K-NN), neural network (NN), and naive Bayes. A K-fold cross-validation technique was applied for validation because it is appropriate for small input data and provides minimum bias during the training process [45]. For this study, the dataset was split into 10 folds (K = 10). The classifiers used for evaluating results are described below.

1) K-NEAREST NEIGHBORS

K-NN is a classification method for categorizing data by using the principle of comparing the data of interest with other information to see how similar it is. This method does not use training data to create a model but instead uses this data as a model. This technique decides which data class is similar or close to by examining certain numbers (K). It is suitable for numerical data to find a way to measure the distance of each attribute in the dataset. K-NN starts by calculating the distance between a new point and each point in the training set to find the closest point to the new point. The obtained distances are sorted, and then the nearest neighbors are found based on the K^{th} minimum distance to predict the class.

2) NEURAL NETWORK

A NN is a mathematical model that is inspired by the structure and functional aspects of biological neural networks. NN comprises an interconnected group of artificial neurons, and it processes information using a connectionist approach to computation. This method can learn a model by a feed-forward NN trained by a backpropagation algorithm (multi-layer perceptron [MLP]). A feed-forward NN indicates connections between the units that do not form a directed cycle, in which the information forwards from the input nodes through the hidden nodes to the output nodes. The MLP principle is an operation of each layer of the hidden layer that has a function to calculate when it receives a signal output from a node in the previous layer, called the activation function. The hidden layer has an important function to convert the data coming into the layer in order to distinguish it using a linearly separable line before sending the data to the output layer. The difference between the actual output and the target output is calculated by propagating forward to notice the error. If the error is large, the parameters (weights and biases) will be updated by propagating backward until the error becomes minimum. Then, the model will make a prediction.

3) NAIVE BAYES

Naive Bayes is a machine learning method built on the principle of probability. This method is used to analyze the probability relying on the Bayesian probability theory. It represents the relationship of conditional probabilities of statistical quantities that determines the probability of a class given some observed attributes presented as:

$$P(i|j) = \frac{P(j|i) \times P(i)}{P(j)}$$

where $P(i\,|\,j)$ refers to the posterior probability that is preserved to attribute j along with class i values. In contrast, $P(j\,|\,i)$ reverses the relation between the class values of the training data, where class i opposes attribute j. $P(i)$ is the prior probability of class i. Moreover, value j denotes $j_1, j_2, \ldots \cap j_n$, whereas n is the attributes of each training data.

IV. RESULTS AND DISCUSSION

Results of attribute selections and comparison results between the proposed IGD method and classical attribute selection methods are presented and discussed in this section.

A. RESULTS OF ATTRIBUTE SELECTION WITH CLASSICAL ATTRIBUTE SELECTION METHODS

Table 5 demonstrates the results of attribute selection using classical attribute selection methods, including information gain, Gini index, chi-squared, and decision tree. This table presents the selected and eliminated attributes based on the whole dataset. From the table, the attributes selected by using information gain and Gini index were TC, LDL-C, HDL-C, and BMI, while SM was eliminated. The attributes selected

TABLE 5. Attribute selection results using the proposed IGD method.

Method	Selected attribute (Relevant attribute)	Eliminated attribute (Irrelevant attribute)
Information Gain	TC, LDL-C, HDL-C, BMI	SM
Gini Index	TC, LDL-C, HDL-C, BMI	SM
Chi-squared	TC, LDL-C, HDL-C, BMI, SM	-
Decision Tree	TC, LDL-C, HDL-C	BMI, SM

by using chi-squared included all attributes. The attributes selected by decision tree were TC, LDL-C, and HDL-C, while BMI and SM were eliminated. The selected attribute represents the relevant attribute that can affect the classification, whereas the eliminated attribute represents the irrelevant attribute that does not affect the classification.

B. RESULTS OF ATTRIBUTE SELECTION WITH IGD METHOD

Table 6 demonstrates the results of attribute selection using the proposed IGD method. In this table, the selected attribute is indicated by the value of normalized information gain that is more than zero ($Z_j > 0$). The eliminated attribute is indicated by the value of normalized information gain that is equal to zero ($Z_j = 0$). For example, the values of normalized information gain of TC, LDL-C, HDL-C, BMI, and SM for elderly person no. 3 were 2.14, 2.07, 0.00, 1.63, and 0.00, respectively. Accordingly, TC, LDL-C, and BMI were selected as the relevant attributes that can affect the classification of elderly person no. 3. At the same time, HDL-C and SM were eliminated, representing the irrelevant attributes that do not affect the classification of elderly person no. 3. The values of normalized information gain of TC, LDL-C, HDL-C, BMI, and SM for elderly person no. 4 were 0.00, 1.47, 0.00, 1.12, and 0.00, respectively. LDL-C and BMI were selected as the relevant attributes that can affect the classification of elderly person no. 4, while TC, HDL-C, and SM were eliminated, representing the irrelevant attributes that do not affect the classification of elderly person no. 4. From this result, it can be recognized that each individual has different selected and eliminated attributes because of diverse health problems for individuals.

C. COMPARISON RESULTS

This section presents comparison results between the proposed IGD method and classical attribute selection methods. Attribute selection is performed in order to select the appropriate attributes according to different feature selection method, which is used for both training and testing set. The comparison results for those methods were performed by applying the selected attributes for classifying groups of elderly people with hypertension, using K-NN, NN, and naive Bayes classifiers. The results of the proposed IGD method and classical attribute selection methods were compared by average classification accuracy using the existing classifiers, as shown in Table 7.

In Table 7, accuracy for elderly group classification employing the relevant attributes selected by information gain and Gini index method provided 59.11% for K-NN, 64.98% for NN, and 64.88% for naive Bayes classifiers. The classification accuracy for information gain and Gini index method was equal because the relevant attributes selected by both methods were the same. In addition, the relevant attributes selected by chi-squared and decision tree method for elderly group classification provided accuracy of 59.53% and 59.01% for K-NN, 64.88% and 64.98% for NN, and 63.44% and 64.57% for naive Bayes, respectively. For elderly group classification employing the relevant attributes selected by the proposed IGD method provided accuracy of 98.76% for K-NN, 98.66% for NN, and 98.76% for naive Bayes.

To compare the classification accuracy results, overall classification performance was explained based on each classifier as follows.

1) K-NEAREST NEIGHBORS

An overall classification performance applying the relevant attributes selected with classical attribute selection methods demonstrates that half of the elderly people were incorrectly classified based on the 5th minimum distance. It was caused by selecting the relevant attributes from the whole dataset, including all attributes of all elderly people, for training, although some attributes were already eliminated. Additionally, some relevant attributes may not apply to some individuals, representing irrelevant attributes for that individual, but those attributes were considered to calculate the distance for all elderly people during the classification process. Using many irrelevant attributes to calculate the distance between a target person for identifying the elderly group and the observed elderly people in the training set can cause poor classification performance. In contrast, based on considering individual attributes, most elderly people were correctly classified. K-NN used only individual relevant attributes to calculate the distance of each elderly person for finding the nearest neighbors based on the 5th minimum distance to classify the elderly group.

2) NEURAL NETWORK

Based on selecting the relevant attributes from the whole dataset, an overall classification performance indicated that almost half of the elderly people were incorrectly classified. It was caused by some relevant attributes selected with classical attribute selection methods that may not apply to some individuals, but those attributes were considered to be input values for training the model. The relevant attributes selected from the whole dataset, including all attributes of all elderly people, and weights with some random values for each attribute were applied to train the model using a backpropagation algorithm. The classification result provided poor accuracy because elderly people have different individual relevant

TABLE 6. Attribute selection results using classical attribute selection methods.

Elderly no.	Results						
	Normalized information gain					*Selected attribute with $Z_j > 0$ (Relevant attribute)*	*Eliminated attribute with $Z_j = 0$ (Irrelevant attribute)*
	TC	*LDL-C*	*HDL-C*	*BMI*	*SM*		
1	2.87	2.01	2.94	2.36	0.00	TC, LDL-C, HDL-C, BMI	SM
2	2.14	2.11	0.00	1.67	0.00	TC, LDL-C, BMI	HDL-C, SM
3	2.14	2.07	0.00	1.63	0.00	TC, LDL-C, BMI	HDL-C, SM
4	0.00	1.47	0.00	1.12	0.00	LDL-C, BMI	TC, HDL-C, SM
5	0.00	0.00	1.65	1.63	0.00	HDL-C, BMI	TC, LDL-C, SM
6	0.00	0.00	1.05	0.00	0.00	HDL-C	TC, LDL-C, BMI, SM
7	0.00	1.25	0.00	0.00	0.00	LDL-C	TC, HDL-C, BMI, SM
8	0.00	0.00	0.00	0.00	0.00	-	TC, LDL-C, HDL-C, BMI, SM
9	0.00	0.00	1.48	1.48	0.00	HDL-C, BMI	TC, LDL-C, SM
10	0.00	0.00	0.00	1.24	1.15	BMI, SM	TC, LDL-C, HDL-C
…	…	…	…	…	…	…	…
971	0.00	1.95	1.99	1.91	0.00	LDL-C, HDL-C, BMI	TC, SM

TABLE 7. Comparison results between the proposed IGD method and classical attribute selection methods.

Method	Classifier		
	K-Nearest Neighbors	Neural Network	Naive Bayes
Information Gain	59.11%	64.98%	64.88%
Gini Index	59.11%	64.98%	64.88%
Chi-squared	59.53%	64.88%	63.44%
Decision Tree	59.01%	64.98%	64.57%
Proposed IGD	98.76%	98.66%	98.76%

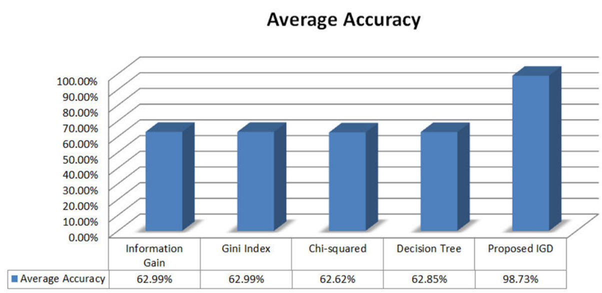

FIGURE 6. Average classification accuracy for each attribute selection method.

and irrelevant attributes. Although the updated weights used to multiply with attributes were adjusted to fit with those input data, the result still provided unsatisfactory classification performance because of diverse individual health conditions.

In contrast, based on considering individual attributes, most elderly people were correctly classified because the proposed IGD method eliminated some individual attributes indicating the irrelevant attributes for developing hypertension levels by transforming its value to zero. The classification result provided good accuracy because the irrelevant attributes for each elderly person were represented as a zero value, which does not affect the classification process. Moreover, the updated weights were adjusted to fit with those attributes.

3) NAIVE BAYES

Overall performance of elderly group classification using a naive Bayes classifier provided half of the elderly people that were incorrectly classified. The relevant attributes obtained from classical attribute selection methods, based on the whole dataset, including all attributes of all elderly people, were applied to classify elderly groups. Some selected attributes may not apply to some individuals, but those attributes were used to find the probability of a target group for all people, which results in poor accuracy. In contrast, using the proposed IGD method to select the relevant attributes based on individual attributes revealed that most elderly people were correctly classified. This satisfying result was obtained by eliminating irrelevant attributes of each elderly person by transforming its value to be zero, which did not affect the classification.

To compare the classification performance, the average classification accuracy for each attribute selection method is presented in Figure 6. Overall, the average classification accuracy of data with the IGD attribute selection method is better than those with classical attribute selection methods. The comparison results confirm that the proposed IGD attribute selection method can improve classification performance, particularly for the case study of elderly people with hypertension. The IGD method selects the individual attributes that are different among individuals, while the classical attribute selection method selects the attributes that are common for all individuals. From this study, the individual attribute's point of view thus presents its potential for use in other medical classification problems for future study.

To summarize, the proposed IGD method was inspired by the concept of a general diagnostic principle of healthcare professionals that considers some individual factors of patients. This concept led to generate the IGD method

TABLE 8. Table of symbols for the proposed IGD method.

Symbol	Description	Symbol	Description
Ep_+	Value of equal proportion for a positive class	k	Significant value for identifying target class
Ep_-	Value of equal proportion for a negative class	pAB_j	Proportion of the acceptable distance
Nj	Total number of attributes	pAC_j	Proportion of the current distance
Ng	Total number of possible target classes	$E(G_j)$	Entropy of each attribute
fg_+	Positive class fraction	$E(G_a)$	Entropy of all attributes
fg_-	Negative class fraction	$Gain(G,j)$	Information gain of the target class regarding all attributes
$E(G)$	Entropy of possible target classes regarding all attributes	Z_j	Normalized information gain of each attribute
dAB_j	Acceptable distance of each attribute	F_j	The j attribute
dAC_j	Current distance of each attribute	et	An eliminated attribute
A_j	Expected level of j attribute	st	A selected attribute
B_j	Acceptable level of j attribute	pAB_{j+}	Proportion of the acceptable distance for a positive class
C_j	Current level of j attribute	pAC_{j-}	Proportion of the current distance for a negative class
K	Actual value of the target		

for selecting individual attributes and eliminating irrelevant attributes for each individual. The proposed IGD method works well with a small dataset for selecting individual attributes representing the actual effect of attributes toward the classification for each individual. However, it is also worth testing the proposed IGD methods with medical classification problems with a larger data set in the future study.

V. CONCLUSION

This study proposes IGD as a new attribute selection method for selecting individual attributes for achieving better classification performance. The experiment is conducted with the group classification of elderly people with hypertension. In this study, classical attribute selection methods, including information gain, Gini index, chi-squared, and decision tree, were employed for selecting relevant attributes. For performance evaluation, the secondary data for 971 elderly people with hypertension were collected. The classification results of data with the IGD attribute selection method and classical attribute selection methods were compared using K-NN, NN, and naive Bayes classifiers. Classification with the IGD attribute selection method provided an average accuracy of 98.73%, which was higher than that of classical attribute selection methods.

APPENDIX A

See Table 8.

ACKNOWLEDGMENT

The authors would like to thank Mae Fah Luang University and CTIF Global Capsule, Aarhus University, for their joint support. Moreover, they thank the hospitals in Chiang Rai, Thailand, for providing the secondary data used in this study.

REFERENCES

[1] W. Xing and Y. Bei, "Medical health big data classification based on KNN classification algorithm," *IEEE Access*, vol. 8, pp. 28808–28819, Nov. 2020.
[2] S. Yang, X. Zhu, L. Zhang, L. Wang, and X. Wang, "Classification and prediction of tibetan medical syndrome based on the improved BP neural network," *IEEE Access*, vol. 8, pp. 31114–31125, Feb. 2020.
[3] S. Balasubramanian and P. Marichamy, "An efficient medical data classification using oppositional fruit fly optimization and modified kernel ridge regression algorithm," *J. Ambient Intell. Humanized Comput.*, vol. 12, no. 3, pp. 3889–3899, Mar. 2021.
[4] J. Helbostad, B. Vereijken, C. Becker, C. Todd, K. Taraldsen, M. Pijnappels, K. Aminian, and S. Mellone, "Mobile health applications to promote active and healthy ageing," *Sensors*, vol. 17, no. 3, p. 622, Mar. 2017.
[5] K. K. F. Tsoi, B. Yip, D. W. H. Au, Y.-H. Kuo, S. Y. S. Wong, J. Woo, and H. M. L. Meng, "Blood pressure monitoring on the cloud system in elderly community centres: A data capturing platform for application research in public health," in *Proc. 7th Int. Conf. Cloud Comput. Big Data (CCBD)*, Macau, China, Nov. 2016, pp. 312–315.
[6] M.-J. Tormo, "Validation of self diagnosis of high blood pressure in a sample of the Spanish EPIC cohort: Overall agreement and predictive values," *J. Epidemiol. Community Health*, vol. 54, no. 3, pp. 221–226, Mar. 2000.
[7] R. D. Giorno, P. S. Heiniger, L. Balestra, and L. Gabutti, "Blood pressure measurements with different currently available methods in elderly hypertensive hospitalized patients: A real world cross-sectional study," *BioMed Res. Int.*, vol. 2019, pp. 1–10, Apr. 2019.
[8] M. Dianatinasab, M. Fararouei, M. Mohammadianpanah, and M. Zare-Bandamiri, "Impact of social and clinical factors on diagnostic delay of breast cancer: A cross-sectional study," *Medicine*, vol. 95, no. 38, p. e4704, Sep. 2016.
[9] P. Croskerry, "Diagnostic failure: A cognitive and affective approach," *Adv. Patient Saf., Res. Implement.*, vol. 2, pp. 241–254. Jan. 2005.
[10] K. Bruchmüller and T. D. Meyer, "Diagnostically irrelevant information can affect the likelihood of a diagnosis of bipolar disorder," *J. Affect. Disorders*, vol. 116, nos. 1–2, pp. 148–151, Jul. 2009.
[11] S. Kjeldsen, R. D. Feldman, L. Lisheng, J.-J. Mourad, C.-E. Chiang, W. Zhang, Z. Wu, W. Li, and B. Williams, "Updated national and international hypertension guidelines: A review of current recommendations," *Drugs*, vol. 74, no. 17, pp. 2033–2051, Oct. 2014.
[12] T. Unger, C. Borghi, F. Charchar, N. A. Khan, N. R. Poulter, D. Prabhakaran, A. Ramirez, M. Schlaich, G. S. Stergiou, M. Tomaszewski, R. D. Wainford, B. Williams, and E. Aletta, "International society of hypertension global hypertension practice guidelines," *Hypertension*, vol. 75, no. 6, pp. 1334–1357, May 2020.
[13] A. Mandal. *Obesity and Blood Pressure*. News-Medical. Accessed: Nov. 3, 2020. [Online]. Available: https://www.news-medical.net/health/Obesity-and-Blood-Pressure.aspx
[14] W. Rogers. (2020). *Mobile App Aims to Help Older Adults Manage Medications*. Kinesiology and Community Health, University of Illinois, College of Applied Health Sciences. Accessed: Nov. 3, 2020. [Online]. Available: https://ahs.illinois.edu/blog/mobile-app-aims-help-older-adults-manage-medications

[15] R. Debon, E. A. Bellei, D. Biduski, S. S. Volpi, A. L. S. Alves, M. R. Portella, and A. C. B. De Marchi, "Effects of using a mobile health application on the health conditions of patients with arterial hypertension: A pilot trial in the context of Brazil's family health strategy," *Sci. Rep.*, vol. 10, no. 1, pp. 1–10, Dec. 2020.

[16] *What is the Place of Lifestyle Changes in the Treatment of Hypertension?*. 2017. Medicographia. Accessed: Nov. 1, 2020. [Online]. Available: https://www.medicographia.com/2017/11/what-isthe-place-of-lifestyle-changes-in-the-treatment-of-hypertension/

[17] S. Chaising and P. Temdee, "Determining recommendations for preventing elderly people from cardiovascular disease complication using objective distance," in *Proc. Global Wireless Summit (GWS)*, Chiang Rai, Thailand, Nov. 2018, pp. 151–155.

[18] S. Chaising, R. Prasad, and P. Temdee, "Personalized recommendation method for preventing elderly people from cardiovascular disease complication using integrated objective distance," *Wireless Pers. Commun.*, vol. 117, pp. 215–233, Mar. 2021.

[19] Z. Deng, Z. Zheng, D. Deng, T. Wang, Y. He, and D. Zhang, "Feature selection for multi-label learning based on F-neighborhood rough sets," *IEEE Access*, vol. 8, pp. 39678–39688, Feb. 2020.

[20] B. Subanya and R. Rajalaxmi, "A novel feature selection algorithm for heart disease classification," *Int. J. Comput. Intell. Inform.*, vol. 4, no. 2, pp. 117–124, Jul./Sep. 2014.

[21] A. Khemphila and V. Boonjing, "Heart disease classification using neural network and feature selection," in *Proc. 21st Int. Conf. Syst. Eng.*, Las Vegas, NV, USA, Aug. 2011, pp. 406–409.

[22] M. A. Jabbar, B. L. Deekshatulu, and P. Chandra, "Prediction of heart disease using random forest and feature subset selection," in *Innovations in Bio-Inspired Computing and Applications*. Cham, Switzerland: Springer, Dec. 2015, pp. 187–196.

[23] A. H. Shahid, M. P. Singh, B. Roy, and A. Aadarsh, "Coronary artery disease diagnosis using feature selection based hybrid extreme learning machine," in *Proc. 3rd Int. Conf. Inf. Comput. Technol. (ICICT)*, San Jose, CA, USA, Mar. 2020, pp. 341–346.

[24] F. Ghasemi, B. S. Neysiani, and N. Nematbakhsh, "Feature selection in pre-diagnosis heart coronary artery disease detection: A heuristic approach for feature selection based on information gain ratio and gini index," in *Proc. 6th Int. Conf. Web Res. (ICWR)*, Tehran, Iran, Apr. 2020, pp. 27–32.

[25] A. N. Syapariyah, A. Saifudin, and T. Desyani, "Feature selection techniques to choose the best features for Parkinsons disease predictions based on decision tree," *J. Phys., Conf. Ser.*, vol. 1477, no. 3, Mar. 2020, Art. no. 032008.

[26] S. V. Pede, S. Chandurkar, and S. Bansode, "Attribute selection to improve accuracy of classification," *Int. J. Comput. Appl.*, vol. 173, no. 5, pp. 975–8887, Sep. 2017.

[27] I. Guyon and A. Elisseeff, "An introduction to feature extraction," in *Feature Extraction*. Berlin, Germany: Springer, 2006, pp. 1–25.

[28] V. Bolón-Canedo, N. Sánchez-Maroño, and A. Alonso-Betanzos, "Feature selection for high-dimensional data," *Prog. Artif. Intell.*, vol. 5, no. 2, pp. 65–75, Feb. 2016.

[29] B. Cancela, V. Bolón-Canedo, A. Alonso-Betanzos, and J. Gama, "A scalable saliency-based feature selection method with instance-level information," *Knowl.-Based Syst.*, vol. 192, Mar. 2020, Art. no. 105326.

[30] R. B. Pereira, A. P. D. Carvalho, B. Zadrozny, and L. H. D. C. Merschmann, "Information gain feature selection for multi-label classification," *J. Inf. Data Manage.*, vol. 6, no. 1, pp. 48–58, Feb. 2015.

[31] Kurniabudi, D. Stiawan, Darmawijoyo, M. Y. Bin Idris, A. M. Bamhdi, and R. Budiarto, "CICIDS-2017 dataset feature analysis with information gain for anomaly detection," *IEEE Access*, vol. 8, pp. 132911–132921, 2020, doi: 10.1109/ACCESS.2020.3009843.

[32] L. H. Patil and M. Atique, "A novel feature selection based on information gain using WordNet," in *Proc. Sci. Inf. Conf.*, London, U.K., Oct. 2013, pp. 625–629.

[33] Y. A. N. G. Ying, J. X. Li, J. C. Chen, C. A. O. Jie, X. F. Lu, S. F. Chen, X. G. Wu, X. F. Duan, X. B. Mo, and D. F. Gu, "Effect of elevated total cholesterol level and hypertension on the risk of fatal cardiovascular disease: A cohort study of Chinese steelworkers," *Chin. Med. J.*, vol. 124, no. 22, pp. 3702–3706, Feb. 2011.

[34] Harvard Health Publishing. (2013). *Blood Pressure High? Control LDL*. Accessed: Nov. 3, 2020. [Online]. Available: https://www.health.harvard.edu/heart-health/blood-pressure-high-control-ldl

[35] A. Israel and E. Grossman, "Elevated high-density lipoprotein cholesterol is associated with hyponatremia in hypertensive patients," *Amer. J. Med.*, vol. 130, no. 11, p. 1324-e7, Nov. 2017.

[36] WHO Regional Office for Europe. (2019). *Body Mass Index—BMI*. Accessed: Nov. 10, 2020. [Online]. Available: https://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi

[37] W. E. Consultation, "Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies," *Lancet*, vol. 363, pp. 157–163, Jan. 2004.

[38] J. E. Winter, R. J. MacInnis, and C. A. Nowson, "The influence of age on the BMI and all-cause mortality association: A meta-analysis," *J. Nutrition, Health Aging*, vol. 21, no. 10, pp. 1254–1258, Dec. 2017.

[39] J. E. Winter, R. J. MacInnis, N. Wattanapenpaiboon, and C. A. Nowson, "BMI and all-cause mortality in older adults: A meta-analysis," *Amer. J. Clin. Nutrition*, vol. 99, no. 4, pp. 875–890, Apr. 2014.

[40] H. Tolonen, H. Wolf, D. Jakovljevic, and K. Kuulasmaa. (2002). *Smoking: Definitions, in: Review of Surveys for Risk Factors of Major Chronic Diseases and Comparability of the Results*. European Health Risk Monitoring (EHRM) Project. Accessed: Nov. 5, 2020. [Online]. Available: https://www.thl.fi/publications /ehrm/product1/title.htm

[41] S. Chaising, P. Temdee, and R. Prasad, "Weighted objective distance for the classification of elderly people with hypertension," *Knowl.-Based Syst.*, vol. 210, Dec. 2020, Art. no. 106441.

[42] A. Alsarah, O. Alsara, and G. Bachauwa, "Hypertension management in the elderly: What is the optimal target blood pressure?" *Heart Views, Official J. Gulf Heart Assoc.*, vol. 20, no. 1, pp. 11–16, Jan. 2019.

[43] P. Liu and J.-G. Zheng, "Blood pressure targets in the hypertensive elderly," *Chin. Med. J.*, vol. 130, no. 16, pp. 1968–1972, Aug. 2017.

[44] A. Alsarah, O. Alsara, and G. Bachuwa, "Hypertension management in older adults," *Avicenna J. Med.*, vol. 8, no. 2, pp. 65–66, Apr. 2018.

[45] J. Brownlee. (Aug. 3, 2020). *A Gentle Introduction to K-Fold Cross-Validation*. Machine Learning Mastery. Accessed: Nov. 10, 2020. [Online]. Available: https://machinelearningmastery.com/k-fold-crossvalidation/

SUPANSA CHAISING received the bachelor's degree in accounting and the master's degree in computer engineering from Mae Fah Luang University, Thailand, in 2014 and 2018, respectively, where she is currently pursuing the Ph.D. degree in computer engineering. Her research interests include artificial intelligence, machine learning, and data analysis.

PUNNARUMOL TEMDEE received the bachelor's degree in electronic and telecommunication engineering, the master's degree in electrical engineering, and the Ph.D. degree in electrical and computer engineering from the King Mongkut's University of Technology Thonburi, Thailand. She is currently an Associate Professor with Mae Fah Luang University, Chiang Rai, Thailand. Her research interests include artificial intelligence and its applications, context-aware computing, and ubiquitous computing.

RAMJEE PRASAD (Life Fellow, IEEE) is currently a Professor of Future Technologies for Business Ecosystem Innovation (FT4BI) with the Department of Business Development and Technology, Aarhus University, Denmark. He is also the Founder President of the CTIF Global Capsule (CGC). He has published over 1000 technical articles, contributed to several books, and has authored or coauthored and edited over 30 books, and holds more than 15 patents. He is also the Founder Chairman of the Global ICT Standardisation Forum for India, established in 2009.