Using Hierarchical Likelihood Towards Support Vector Machine: Theory and Its Application

REZZY EKO CARAKA[1,2,3], **YOUNGJO LEE**[1,2], **RUNG-CHING CHEN**[3], **(Member, IEEE),**
AND TONI TOHARUDIN[4]

[1]Laboratory of Hierarchical Likelihood, Research for Basic Sciences, College of Natural Sciences, Seoul National University, Seoul 08826, South Korea
[2]Department of Statistics, College of Natural Sciences, Seoul National University, Seoul 08826, South Korea
[3]Department of Information Management, College of Informatics, Chaoyang University of Technology, Taichung City 41349, Taiwan
[4]Department of Statistics, Faculty of Mathematics and Natural Science, Universitas Padjadjaran, Bandung 45363, Indonesia

Corresponding authors: Rezzy Eko Caraka (rezzy94@snu.ac.kr), Youngjo Lee (youngjo@snu.ac.kr), and Rung-Ching Chen
(crching@cyut.edu.tw)

ABSTRACT The H-likelihood method proposed by Lee and Nelder (1996) is extensively used in a wide range of data. In terms of direction, repetitive measured data within classification can be examined employing hierarchical generalized linear models (HGLMs). Whether we are concerned in multiple endpoints which are correlated, instead Multivariate Double Hierarchical Generalized Linear Models (DHGLM) can be taken into consideration. This article addresses the implementation of this principle to vector selection and support machines. Based on the analysis with the fish morphology class *Sardinella lemuru* (*Bali sardinella*) and setting the best epsilon 0.7 cost 4 parameter reaching best performance: 0.2327401. Predictive value of fish sex was calculated 0.997319 and Region under the curve: 0.8967. At the same time, we extend the large-scale case studies for stress testing of the SVM method by using three datasets from UCI machine learning repository including the bank marketing dataset, the car evaluation database and human activity recognition using smartphones dataset. In a nutshell by employing SVM-DHGLM increased the accuracy, precision, recall, for feature selection and classification. Long story short, the *H*-likelihood provides an excellent and usable structure for statistical inference of the unobservable general deterministic model, while preserving the advantages of the original probability structure for fixed parameters. We presume that more new groups of models will be created and that the *H*-likelihood will be commonly used for their inferences and the application in big data and machine learning.

INDEX TERMS Hierarchical, likelihood, SVM, DHGLM, classification.

I. INTRODUCTION

Data mining, also referred to as knowledge discovery in databases (KDDs), is a procedure that involves processing as well as using historical data to identify patterns and relationships, associations or relationships in massive amounts of data [1]. Data mining implementation could be used to support future decision-making [2]. Machine learning is a branch of artificial intelligence or data analytics that deals with the development of algorithms that can be configured to learn from previous information. Machine learning is indeed a computational method for data mining. Data mining

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

required to learn processes are frequently divided into two main types, respectively supervised and unsupervised [3]. During the first, unsupervised learning approach, the learning method is applied without any training stage without any target or learner, throughout this case the data label [4].

There are already several learning algorithms that have already been established, which include K-Nearest Neighbor [5], Artificial Neural Network [6], Naïve Bayes [7], Support Vector Machine [8] and many more [9]. This algorithm has strengths and weaknesses [4]. However, all implementations have had the same concept, in other words learning because then, at the end of the training, the model can clearly identify - input data mostly on output class label. In this method, to properly assess the decision function, the

separator function or indeed the regression function, the data which has labels or goals is used to learn. Forward to training with sufficient data, machine learning, instance SVM, could be used to predict or identify what decisions have been made if new data is entered and also the result is unspecified [10]. Unless the performance either at time of testing by itself doesn't perform as expected, the parameters of the machine learning function can be modified to boost the accuracy of the model. Throughout the case of SVM, the parameters almost always configured are cost, epsilon, gamma parameters [11], [12]. Forecasting or predictive processes in data mining can be classified into two major groups, respectively classification and ability to apply knowledge or regression [13]. In classification, the output of each data is an integer or discrete number, while being in regression, the output of each data has been said to be continuous [14].

The construction of a classification model in SVM is based on risk minimization, which contributes throughout the ability to explain problems well though and overcome overfitting [15]. Including its generalizability, SVM is capable of producing high precision and a comparatively modest error rate. In its development, SVM has been successfully used to solve problems in various fields, including hyperspectral data [16], Alzheimer's disease [17], [18], thyroid disease [19]. Besides in ecology [20], including coral reef fish detection and recognition in underwater videos [21] and remote sensing [22]. The ability of SVM as a classification method can be compared with other classification methods [23].

The SVM approach has some limitations, particularly regarding it has difficulty determining the optimal parameter value. In fact, setting the parameter values correctly will improve the classification accuracy of the SVM model [24]. In order to get the parameters that generate the best Classification models, the optimization of the parameter is performed in the SVM model [25]. Optimization of these parameters means the determination among the most efficient hyper parameter for the SVM model and the implementation of the SVM model with the best classification performance [26]–[28]. The Grid Search method is the most commonly used method of optimizing the process parameters [29], [30].

The consideration of a good and sufficient statistical method for the needs of the study is a basic principle in the conduct of statistical analysis [31], [32]. GLM, which also has a random effect as a predictor, can indeed be modified to HGLM [33]. The whole strategy does indeed have a lot of advantages that GLM doesn't provide [34]. The aim of inferential statistics is to derive from the results the findings of the analysis accompanying the population. HGLM is a statistical modeling that appears to offer consistency in modeling and broad application, including the use of penalized-likelihood scenarios [35], factor analysis ordinal data that can be used for structural equation modeling [36], modeling longitudinal outcomes time to event [37].

Lee and Nelder [38] address inferential statistical models which mostly combine unobserved random variables, with
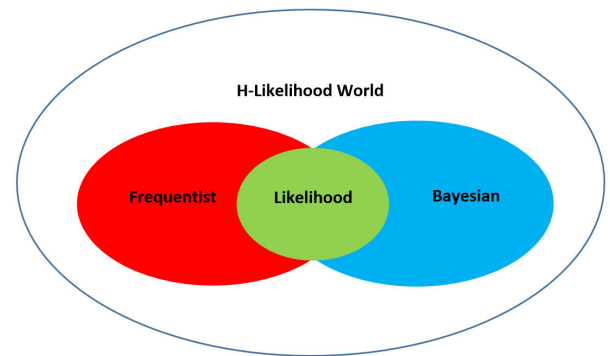

FIGURE 1. The *H*-likelihood world [46].

the exception of output value, missing data, latent variable, component, and potential outcome. Although there are benefits of using *H*-likelihood [39]. With that being said, we can develop fast and efficient computational techniques for model fitting. Furthermore, we can integrate inference statistical models for unobserved samples and then predict the model's output value by including these unobserved variables. This same marginal probability function is used for the inferential fixed effects of both the classical approach and the h-likelihood. Even so, for only certain case studies, the Laplace approximation [40]–[42].

Therefore, this article use H-likelihood to establish a support vector machine by using two case studies. First, the feature selection morphology of *Sardinella lemuru (Bali sardinella)* and the second case by using three datasets from UCI machine learning repository including the Bank marketing dataset, the car evaluation database and Human Activity Recognition Using Smartphones Dataset. Section 2 will address the principle of H likelihood, explicitly and in depth how to derive it, as described in Section 3. In line with this, the actual data set implementation is in Section 4. Finally, the conclusion and possible studies are discussed in Section 5.

II. THE PRINCIPLE OF H-LIKELIHOOD

We aim to have a paradigm shift from the probability-based reasoning to the extended likelihood-based reasoning. Each paradigm constitutes a world (normal science) characterized by basic propositions. Normal science will be very fruitful in describing by deductive reasoning till anomalies (paradoxes) arise.

In modelling techniques random effects perhaps with a flat prior, the *H*-likelihood is equivalent to a Bayesian posterior distribution as well as is not controversial for a Bayesian statistician. Figure 1 represents either Bayesian and frequentist may readily embrace the notion of statistical probability [43], [44]. It enables most Bayesian credible interval interpretations and frequentist confidence interval interpretations. As a result, the *H*-likelihood method attempts to combine the two worlds of frequent statistics and Bayesian statistics, which may be conflicting in most [45].

Hierarchical GLMs (HGLMs) had already formed by [46], [38] which seem to be the GLMs in which the linear predictor includes either fixed and random effects which described as conditional on random effects u, reactions y obey the GLM family of distributions [39] which represents in Eq 1.

$$E(y \mid u) = \mu \quad \text{and} \quad var(y \mid u) = \phi V(\mu) \quad (1)$$

To begin with, we can write the double hierarchical GLM by Eq2 below:

$$
\begin{aligned}
h &= l_1(\beta_1, \gamma_1, \beta_2, \gamma_2, ; y_1, y_2, | v_1, b_1, v_2, b_2) \\
&\quad + l_2(\lambda; v_1, v_2) + l_3(\alpha_1, \alpha_2; b_1, b_2) \\
&= \frac{1}{2} \log |2\pi \Phi_1| \\
&\quad - \frac{1}{2}(y_1 - X_1\beta_1 - Z_1 v_1)^T \Phi_1^{-1}(y_1 - X_1\beta_1 - Z_1 v_1) \\
&\quad - \frac{1}{2} \log |2\pi \Phi_2| \\
&\quad - \frac{1}{2}(y_2 - X_2\beta_2 - Z_2 v_2)^T \Phi_2^{-1}(y_2 - X_2\beta_2 - Z_2 v_2) \\
&\quad - \frac{1}{2} \log |2\pi \Sigma_R| - \frac{1}{2} v^T \Sigma_R^{-1} v \\
&\quad - \frac{m}{2} \log(2\pi \alpha_1) - \frac{1}{2\alpha_1} b_1^T b_1 \\
&\quad - \frac{m}{2} \log(2\pi \alpha_2) - \frac{1}{2\alpha_2} b_2^T b_2 \quad (2)
\end{aligned}
$$

where Φ is a diagonal matrix of $\phi_{ij} = \exp(G_{ij}^T \Upsilon + F_{ij}^T b_i)$ The HGLM model is confirmed by additionally setting various models to the random effect. It can be lengthened. Lee and Nelder set a model for the mean and scatter of random effects Introduced the model of Double HGLM. The multilevel generalization model is Expanded into a three-element dual multi-level generalized linear model. Under the conditions given random effects (a, u), the response variable y satisfies $E(y|a, u) = \mu$ and $var(y|a, u) = \phi V(\mu)$. Under the condition given random effect u, linear prediction function for μ (linear predictor) is given in the form of the following HGLM in Eq3.

$$\eta = g(\mu) = X\beta + Zv \quad (3)$$

The scattering parameter λ for the random effect u has the following HGLM form.

$$
\begin{aligned}
\xi_m &= h_m(\lambda) = G_m \gamma_m \\
\xi &= h(\phi) = G\gamma + Fb \quad (4)
\end{aligned}
$$

Eq4 describes where $h()$ is the link function and G and F are model matrices. Random monotony for number $g_D()$ with $b = g_D(a)$ is a random effect, and γ is a fixed effect. Random effect. The parameter α for a has the form of the following GLMs. The dispersion model for ϕ and φ is in the form of HGLM in Eq5.

$$\xi_D = h_D(\alpha) + G_D \gamma_D \quad (5)$$

where h_D is the GLM connection function and γ_D is the fixed effect. Likelihood functions for dual multi-level generalization models (DHGLMs) can defined in Eq 6.

$$L(v, b, y|\beta, \phi, \lambda, \alpha) = f_{\beta,\phi}(y|v, b) f_\lambda(v) f_\alpha(b) \quad (6)$$

Therefore, Eq7 represents H-likelihood type for inference from dual multi-level generalization models (DHGLMs)

$$h = \log f(y|v, b; \beta, \phi) + \log f(y; \lambda) + \log f(b; \alpha) \quad (7)$$

where function of $f(y|v, b; \beta, \phi)$, function of $f(y; \lambda)$, and function of $f(b; \alpha)$ each (v, b) are given y, u and is the conditional distribution of b respectively. Besides, Eq8 with y express as marginal likelihood $L_{v,b}$ which can be written as follows:

$$
\begin{aligned}
L_{v,b} &= \log \int \exp(h) \, dv db \\
&= \log \int \exp L_v \, db = \log \int \exp L_b \, dv \quad (8)
\end{aligned}
$$

where $L_v = \log \int \exp(h) \, dv$ and $L_b = \log \int \exp(h) \, db$. Then, the marginal likelihood $L_{v,b}$ provides a rational reasoning process for fixed parameters. But,In the case of general reasoning, there is no information about random parameters (v, b) that are not observed. Neverthless, $L_{v,b}$ is not enough. In line with this, Lee and Nelder are the evaluation criteria for DHGLMs (v, β) For h, (b, γ, γ_M) also $p_\beta(L_v)$ and γ_D.

III. NEW FRAMEWORK
A. EMPLOYING IN SUPPORT VECTOR MACHINE
The basic objective of the support vector machine is to find a separating hyperplane that optimizes the link between different classes. It is also established the new information, where it would be included in the predictions of a classification purpose, or this condition may well be called a best generalization. The model created by the SVM depends on only the subset of the training, although the cost function of developing the model does not directly affect the training data further than the margin. Eq9 as one of the separation boundary planes, the hyperplane.

$$\left\{ x : f(x) = x^T \beta + \beta_0 = 0 \right\} \quad (9)$$

where β is weight coefficient vector β_0 as a bias $\|\beta\| = 1$. Besides, Classification criteria centered on this hyperplane $G(x) = sign(x^T \beta + \beta_0)$. Constraint $y_i(x^t \beta + \beta_0) > M, i = 1, 2, \ldots, n$ about $max_{\beta, \beta_0} C$. The distance between the support vector and the optimal hyperplane is $C = \frac{1}{\|\beta\|=}$. Therefore, it can be represented as the largest of the vertical distances between the training vector X_i and the hyperplane is called margin. And the only hyperplane that maximizes the margin $C > 0$ of the training materials. Find it. In other words, to solve the SVM, the following block optimization. It becomes a problem.

$$
\begin{aligned}
&\text{minimize } \beta^t \beta \\
&\text{subject to } y_i(x^t \beta + \beta_0) \geq 1
\end{aligned}
$$

However, the above-mentioned formula is applicable only to linearly separable sets. All. However, it is hard to find hyperplanes that perfectly separate these vectors [12]. Therefore, considering the vectors included in the domain of other groups, the slack variable $\xi_1, \xi_2, \ldots, \xi_n$ this method Produces places constraints like this:

$$min_{\beta, \beta_0} \|\beta\|$$
$$s.t \; \forall i, \; y_i \left(x^t \beta + \beta_0 \right) \geq 1 - \xi_i,$$
$$\xi_i \geq 0, \sum_{i=1}^{N} \xi_i \leq constant$$

ξ Re-expressing the constraints for the convenience of calculation, the expression in the form of consensus.

$$min_{\beta, \beta_0} \frac{1}{2} \|\beta\|^2 + C \sum_{i=1}^{N} \xi_i$$
$$s.t \; y_i \left(x^t \beta + \beta_0 \right) \geq 1 - \xi_i \quad \forall i$$

Eq10 represents C tuning parameter plays the role importance. C points to be considered as the support vector you will get a lot of numbers. Using the Lagrange multiplier β_i, β_0, ξ_i solving the problem about in Eq11.

$$L_p = \frac{1}{2} \|\beta\|^2 + C \sum_{i=1}^{N} \xi_i$$
$$- \sum_{i=1}^{N} \alpha_i \left[y_i \left(x_i \beta + \beta_0 \right) - \left(1 - \xi_i \right) \right] - \sum_{i=1}^{N} \mu_i \xi_i$$
$$\hat{\beta} = \sum_{i=1}^{N} \alpha_i y_i x_i$$
$$0 = \alpha_i y_i$$
$$\hat{\alpha}_i = C - \mu_i, \forall i, \tag{10}$$

Substituting can be expressed as:

$$L_D = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \alpha_i \alpha_i' y_i y_i' x_i^T x_i',$$
$$0 \leq \alpha_i \leq C$$
$$\sum_{i=1}^{N} \alpha_i y_i = 0. \tag{11}$$

Variables to be estimated according to Kuhn-Tucker theorem β_i, β_0, ξ_i be made of From the saddle point of space. Besides Eq12 explains the relationship between Lagrange multiplier and conditional expression is shown below.

$$\alpha_i \left[y_i \left(x_i^T \beta + \beta_0 \right) - \left(1 - \xi_i \right) \right] = 0$$
$$\mu_i \xi_i = 0$$
$$y_i \left(x_i^T \beta + \beta_0 \right) - \left(1 - \xi_i \right) \geq 0 \tag{12}$$

The values satisfying the above conditional expression are the relationship between ξ_i and support vector seen earlier. It is the basis for contentment and a unique solution. Now L_D is a regular convex.

IV. ANALYSIS AND RESULTS

Similar to the SVM estimation equation obtained Eq13 the fixed effect of the Bernoulli distribution and Double H-likelihood with random effect of multivariate normal distribution and chi-square distribution. More instance $v = (v_1, v_2, \ldots, v_{1n})$ and for each point w.

$y_i | v \sim bernoulli(p_i)$ with the logit link,
$v | w \sim MVN$ with $\mu = 0$, precision $= w \times \frac{K}{\lambda}$

$$w \sim \chi^2 \frac{\alpha}{\alpha}$$
$$K = \langle x, x' \rangle$$
$$logit(p_i) = \beta_0 + \sum_{j=1}^{N} K_{ij} v_j,$$
$$h = \log H = \sum_{i=1}^{N} \left[y_i \log \frac{p_i}{1 - p_i} + \log (1 - p_i) \right]$$
$$- \frac{N}{2} \log (2\pi) + \frac{w}{2} \log \left(\log \left(\frac{K}{\lambda} \right) \right) - v^T \frac{wK}{2\lambda} v$$
$$- \log \Gamma \left(\frac{\alpha}{2} \right) - \frac{\alpha}{2} \log 2 + \left(\frac{\alpha}{2} - 1 \right) \log w - \frac{1}{2} w$$
$$p_v(h) = h - \frac{1}{2} \log \left| \left(\frac{\frac{\partial^2 h}{\partial v^2}}{2\pi} \right) \right| \tag{13}$$

Eq13 represents of h is h-log likelihood and $p_v(h) =$ is adjusted profile log likelihood. In addition, λ and α corresponds to the tuning parameter of SVM. In line with this, β_0 and v to maximize h, which estimate w. To run the estimation can be perform by The iterative weighted least squares and Newton-Raphson method are used for estimation. Repeat the process below until convergence. h estimates the w_t, $p_w(h)$ can estimate by $v_{0(t)}$, $p_w(h)$ can estimate by $\beta_{0(t)}, w_t, v_t, \beta_{0(t)}$ Estimate w_{t+1}, v_{t+1} after update, respectively. Also, Estimation of $\beta_{0(t+1)}$ with $p_{w+1, v+1}(h)$.
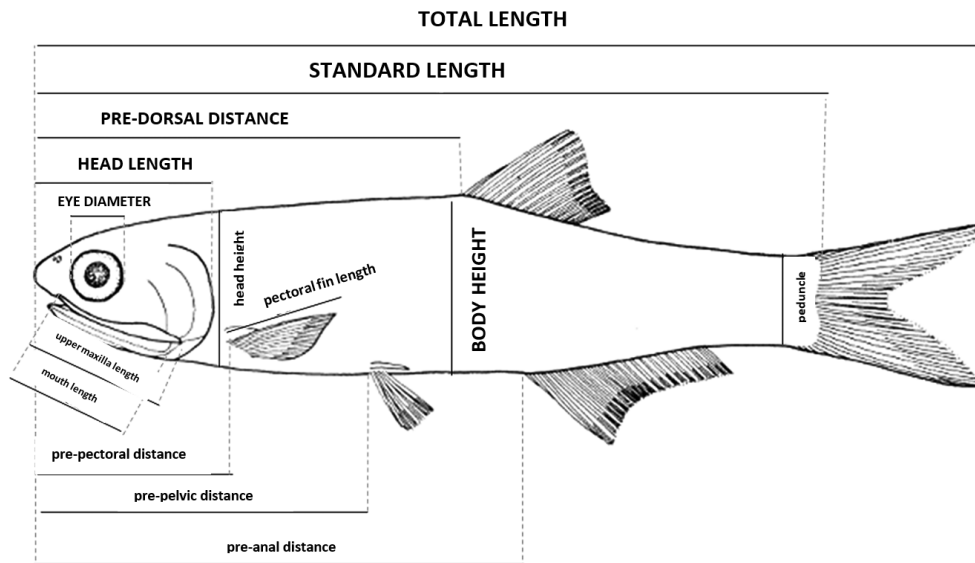
A. EMPLOYING IN SARDINELLA LEMURU (BALI SARDINELLA) DATASET

Sardinella lemuru (Bali sardinella) such as a species of ray-finned fish as in genus of Sardinella discovered throughout the Eastern Indian Ocean as well as the Western Pacific Ocean, inside a neighbourhood which reaches from southern Japan through all the Malaya Archipelago to Western Australia. To begin with, Table 1 and Figure 2 are describes descriptive statistics regarding morphology. the standard deviation is a measure of both the amount of variance or summarizes features of a collection of variables. A small standard deviation indicates that the values tend to be identical to the mean value, while a high standard deviation indicates that the data are distributed over a broader variety. More clearly, total length, fork length, Standard length with sex type contain the differences compares other variables.

The suitability judgment of the model is confirmed through two indicators. First, different Find the sum of the numbers of misclassified colors. Second, distribution of data The probability of the region in error is calculated by using a random

TABLE 1. Summary statistics of all morphology *Sardinella lemuru (Bali Sardinella)*.

Variable	Type	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum	IQR
	Female	18.932	0.154	1.683	16	17.4	18.5	20.4	22	3
Total length (TL) (X₁)	Male	18.322	0.158	1.601	15.2	17	17.9	19.9	21.6	2.9
	Unidentified	18.343	0.117	1.437	15.2	17.4	18.3	19	22.2	1.6
	Female	16.97	0.14	1.523	14.3	15.7	16.5	18.4	20	2.7
	Male	16.447	0.139	1.411	14.2	15.4	16.2	17.8	19.5	2.4
Fork length (X₂)	Unidentified	16.425	0.104	1.279	13.7	15.7	16.4	17	20	1.3
	Female	16.009	0.135	1.478	13.32	14.73	15.74	17.39	18.65	2.66
	Male	15.55	0.144	1.466	13.29	14.33	15.07	16.97	19.9	2.64
Standard length (X₃)	Unidentified	15.405	0.106	1.298	12.9	14.548	15.355	15.93	19.7	1.383
	Female	12.419	0.112	1.217	10.07	11.37	12.09	13.57	15.25	2.2
	Male	12.067	0.113	1.147	10.1	11.16	11.8	13.16	14.25	2
Pre-anal length(X₄)	Unidentified	11.968	0.0829	1.015	10.02	11.248	11.845	12.432	14.42	1.185
	Female	7.2361	0.0638	0.6963	5.9	6.78	7.18	7.79	9.09	1.01
	Male	6.9762	0.0638	0.647	5.38	6.46	6.92	7.51	8.36	1.05
Pre-dorsal length (X₅)	Unidentified	6.9435	0.0518	0.6348	4.27	6.5275	6.965	7.28	8.57	0.7525
	Female	8.3561	0.0814	0.8881	4.21	7.76	8.25	9.07	10.48	1.31
	Male	8.1706	0.0767	0.7784	6.54	7.59	8	8.88	9.52	1.29
Pre-pelvic length(X₆)	Unidentified	8.12	0.0575	0.7039	6.26	7.6775	8.1	8.4925	9.8	0.815
	Female	4.1552	0.0371	0.4051	3.16	3.87	4.17	4.47	5.44	0.6
	Male	4.077	0.0393	0.399	2.71	3.86	4.07	4.38	4.97	0.52
Pre-pectoral length (X₇)	Unidentified	4.0811	0.0353	0.432	3.16	3.8275	4.07	4.2725	6.61	0.445
	Female	3.5059	0.0527	0.5747	1.72	3.17	3.48	3.93	4.74	0.76
	Male	3.4679	0.047	0.4771	2.03	3.14	3.39	3.85	4.58	0.71
Body depth (X₈)	Unidentified	3.4943	0.0414	0.507	1.82	3.19	3.395	3.8225	4.91	0.6325
	Female	4.2266	0.0397	0.4333	2.6	4	4.29	4.51	5.52	0.51
	Male	4.1728	0.0343	0.3485	3.02	3.96	4.18	4.44	4.89	0.48
Head length (HL) (X₉)	Unidentified	4.1637	0.0306	0.3752	2.74	3.975	4.215	4.3625	4.96	0.3875
	Female	0.8505	0.00911	0.09943	0.43	0.8	0.86	0.92	1.07	0.12
	Male	0.83243	0.00902	0.09159	0.48	0.77	0.83	0.9	0.99	0.13
Eye diameter (X₁₀)	Unidentified	0.8394	0.0102	0.1248	0.34	0.8	0.84	0.88	1.87	0.08
	Female	1.5968	0.0139	0.1512	1.05	1.49	1.6	1.7	2.01	0.21
	Male	1.5767	0.0144	0.146	1.13	1.48	1.57	1.68	1.93	0.2
Pre-orbital length (X₁₁)	Unidentified	1.5629	0.0134	0.1641	1.05	1.47	1.565	1.66	2.5	0.19

FIGURE 2. Morphology *Sardinella lemuru (Bali sardinella)*.

sample that knows. These two indicators are named Misclassification and Test error, respectively. Misclassification $= \sum_{i=1}^{N} I_{\hat{y}_i}(y_i)$. Test error $= \int_{y=0} P(\hat{y}_i = 1 | y = 0) P(y)\, dy + \int_{y=1} P(\hat{y}_i = 0 | y = 1) P(y)\, dy$. The smaller the misclassification and test error, the better the model fit. The SVM model using Hinge loss uses the svmpath R package [47]. It was fitted and compared with the model of the DHGLM method. First, the Radial kernel and set each d and tuning parameter the model is estimated by varying the levels of λ and α_a. We now define the conditional Akaike information

TABLE 2. Importance variable.

Variable	Mean Importance	Median Importance	Minimum Importance	Max Importance	Normal Hits	Decision
X_1	17.342386	17.424134	13.3269542	19.597581	1	Confirmed
X_2	13.416093	13.479939	11.4270542	15.458715	1	Confirmed
X_3	9.697339	9.701476	6.9384442	12.3632	1	Confirmed
X_4	6.345838	6.259472	3.6562721	9.905094	0.94949495	Confirmed
X_5	6.875657	6.983328	3.9206249	9.471292	0.96969697	Confirmed
X_6	6.842935	6.795067	4.3526129	9.47106	0.95959596	Confirmed
X_7	4.094106	4.050613	1.9037203	7.564499	0.71717172	Confirmed
X_8	1.55074	1.413015	-0.8465973	4.240001	0.09090909	Rejected
X_9	3.147392	3.065962	1.0620197	6.398725	0.48484848	Tentative
X_{10}	5.882047	5.833342	2.6778448	9.132543	0.93939394	Confirmed
X_{11}	2.376187	2.299845	-0.8877425	5.408358	0.37373737	Tentative

criterion (AICc) for double hierarchical generalized linear model to be used in model selection. Assume the true conditional distribution of y is $f_1(y \mid u, a)$ and that u are the true random effects vector for mean and dispersion with distributions $f_2(u)$ and $f_3(a)$, respectively. The prediction dataset is y^* such that y^* and y are independent conditional on u and a, and from the same distribution $f_1(. \mid u, a)$ In other words, y and y^* share the same random effects u and a, but differ in their error term in Eq 14.

$$
\begin{aligned}
AICc &= -2E_{f(y,u,a)}E_{f(y^* \mid u,a)} \log g\{y^* \mid \hat{\beta}(y), \hat{v}(y), \hat{\gamma}(y), \hat{b}(y) \\
&= \int -2 \log g \left\{y^* \mid \hat{\beta}(y), \hat{v}(y), \hat{\gamma}(y), \hat{b}(y)\right\} \\
&\quad \times f_1(y^* \mid u, a) f(y, u, a) \, dy^* dy du da
\end{aligned} \tag{14}
$$

Assume, the data y are generated from the DHGLM. Then, the AICc is given as follow in Eq15.

$$
AICc = -2 \log g \left(y \mid \hat{\beta}(y), \hat{v}(y), \hat{\gamma}(y), \hat{b}(y)\right) + 2\hat{p}_h \tag{15}
$$

At the same time, Eq16 represents $\hat{p}_h = PD = tr\left(H^{-1}H^*\right)$. H explains as Hessian matrices of H-Likelihood and conditional log likelihood $l_1(\beta, \Upsilon; y \mid v, b)$ with respect to $(\beta, v, \Upsilon, b, \lambda)$. In line with this, conditional log likelihood l_1 does not depend on the dispersion parameter λ.

$$
\begin{aligned}
H^* &= \begin{bmatrix} H^*_{aa} & 0 \\ 0 & 0 \end{bmatrix}; \quad H = \begin{bmatrix} H_{aa} & H_{ab} \\ H_{ba} & H_{bb} \end{bmatrix} \\
&= \begin{pmatrix} -\dfrac{\partial^2 h}{\partial a \partial a^T} & -\dfrac{\partial^2 h}{\partial a \partial b^T} \\ -\dfrac{\partial^2 h}{\partial a \partial b^T} & -\dfrac{\partial^2 h}{\partial a \partial b^T} \end{pmatrix}
\end{aligned} \tag{16}
$$

In this article, using 11 independent variables and prediction of the sex status of fish will be carried out, namely Male (1) and Female (0). In the first stage, feature selection can be performed by finding the mean value for sex as described in Table 2. Fish sexuality is split into different, male and female. Secondary and primary sexual features are used to differentiate between male and female fish. Primary sexual characteristics are characterized by the presence of organs that are directly connected to the reproductive process,

including ovaries and vessels in female fish and testicles and vessels in male fish. Meanwhile, secondary sexual characteristics are useful for distinguishing male and female fish which can be seen from the outside or in terms of morphology. Secondary sexual characteristics can be divided into two parts, namely secondary sexual which is temporary which only appears during the spawning season and secondary sexual which is permanent which occurs at the time before and after spawning. Sex comparisons can be used to predict the success of spawning, namely by looking at the ratio of the number of male and female fish in the waters, it also affects the reproduction, recruitment and conservation of these fish resources.

Features are classified as raw measuring operations and are executed on a group of objects that encourages that classification to be separated from many other categories in the same broad class. Although features may not have been easily accessible in data sets given the high dimension of raw data, two priorities became commonly regarded for both the extracting features in classification tasks. Reducing the dimension in the sequence space might well reduce the size of the original data by selecting the features that still represents the data adequately and precisely.

Rendering yet most effective features for ensuring classification performance the feature selection is necessary and also has a profound effect on classier application. Unless the features indicate significant differences through one class to the, this same classifier can indeed be designed quite quickly and efficiently with performance improvement. In this case Transform based dimensionality reduction uses a transformation function to extract features from raw data. Linear transformations are popular and widely use due to its convenience, whereas nonlinear transformations have increasingly become recommended and examined.

A linear feature extraction method stems from $Y = A.X$, where $Y \epsilon R^n$ and R^n stands for all vectors with n components, each of which is a real number. A is generally a $n \times m$ matrix taking a vector with n components and generating a vector with m components. There are several transformations which are used very commonly. Transformation technique has a certain application domain, including benefits and drawbacks.
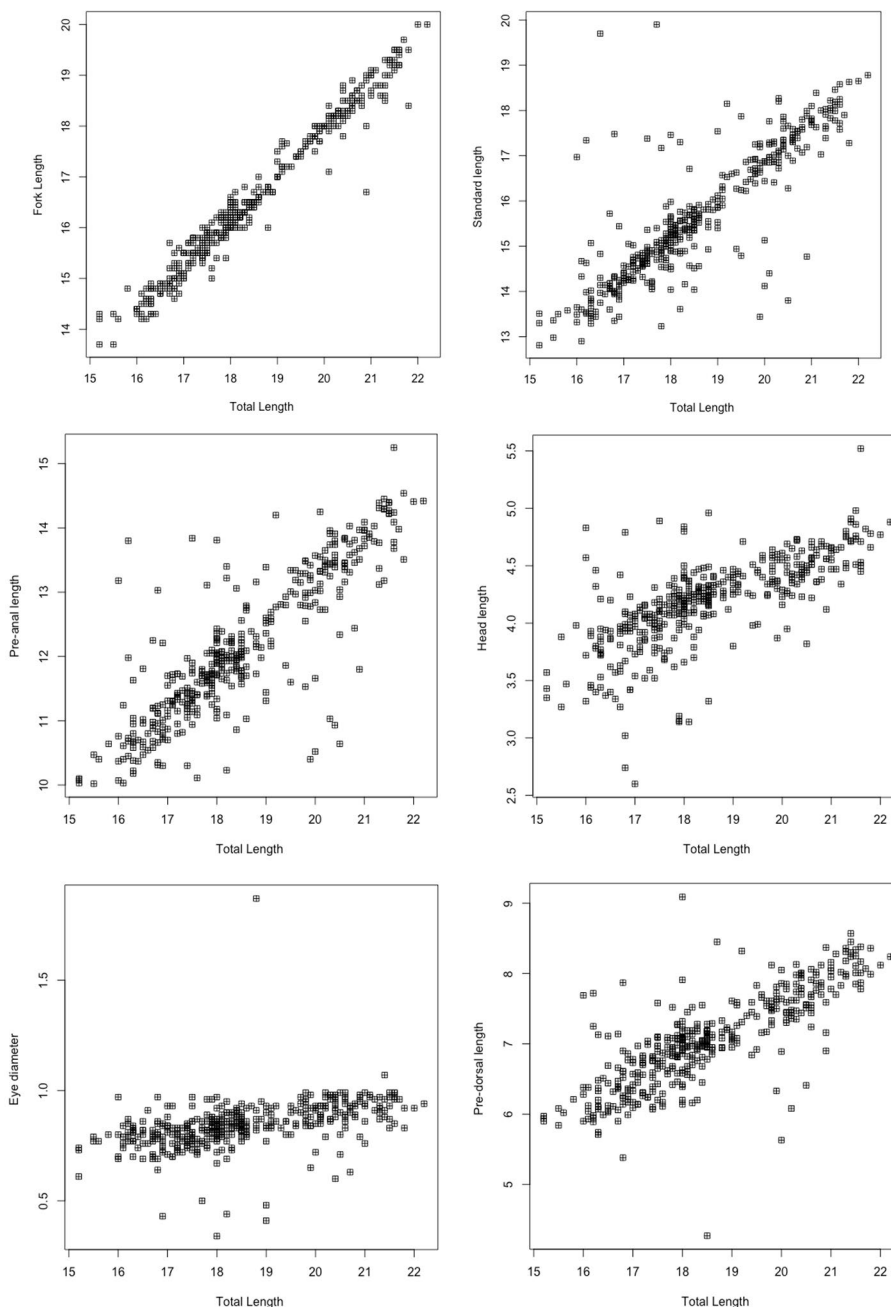
FIGURE 3. Total length with best feature importance.

Transformation functions *A* can usually been divided into two main categories.

In some cases, we may describe transformation on the basis of inhibitory understanding. However, the theories of likelihood at a certain time did not attract sufficient enthusiasm in the statistical community. That principle of likelihood was indeed a result of the much more accepted principles for sufficiency, that even a reasonable fixed comprehension all of the factual information from of the study as well as the circumstance that perhaps the experimental studies never carried

out that are insignificant to its implication. Although since, the concept of likelihood has become one of the fundamental principles of inferential statistics. Possible future methods have been suggested in areas including the calculation of nuisance parameters

Based on Table 2, it can be explained that the only variable that can be used to classify the sex of fish is Body depth because the value is below 0.1. However, the variable Head length and Pre-orbital length are tentative because their values are below 0.50. Thus, the remaining other variables can be
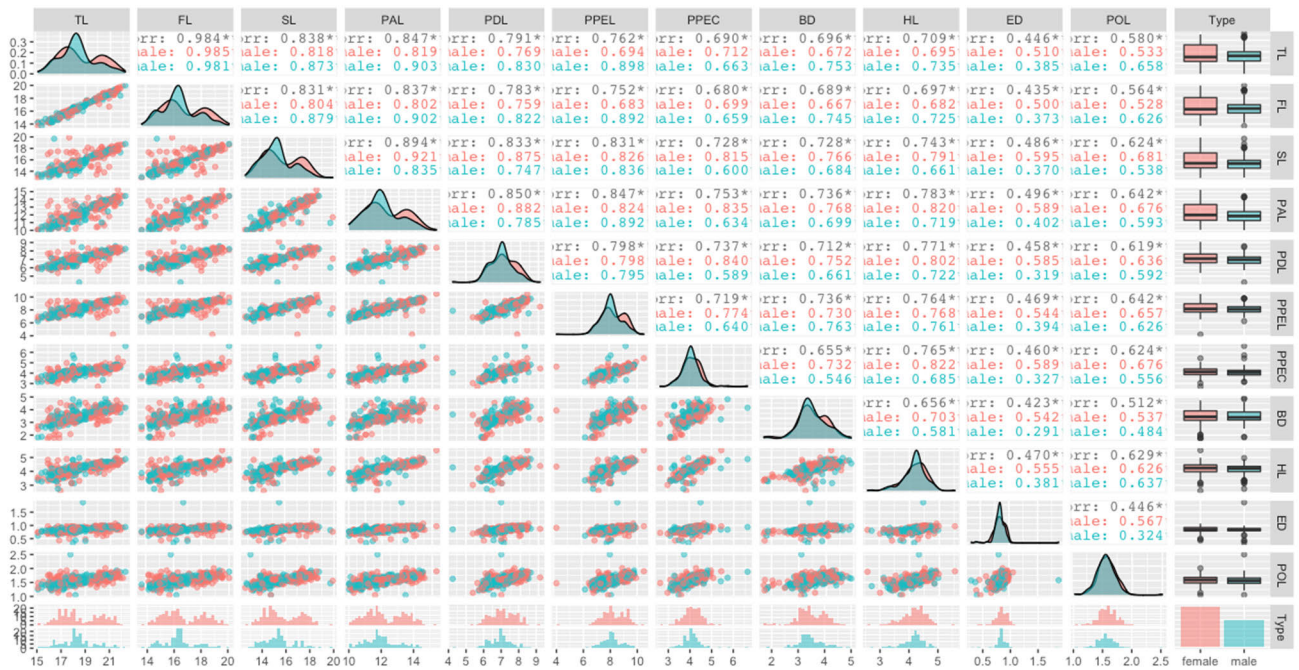
FIGURE 4. Desriptive plot.

used for sex classification of *Sardinella lemuru* (*Bali sardinella*).

From the expert system, it was explained that visually what can distinguish the sex of fish is based on the Total length, Head Length, Eye Diameter, and Fork length so that Figure 3 and Figure 4 explain the relationship of all these variables. The size ranges shown in Table 2 describe the measurement results that did not differ greatly for each sex in the same habitat. This is because the environmental conditions that take place in it are relatively homogeneous, so that this condition does not have a big influence on the size difference between male and female fish.

In addition, the proportion of food availability in the fish cage is relatively secure and has the opportunity to be divided into the same proportion, so that the opportunity to compete is relatively small. It is appropriate that the internal factor affecting fish size is the availability of food.

There are differences in morphometric characters that are also seen in each gender in Table 2, due to differences in age or sex. Environmental conditions in nature tend to fluctuate, especially in terms of competition, so that in order to defend oneself from predators, natural pressure or capture, it forces oneself to grow rapidly. Fish increase body size, especially limb parts such as fins.

This occurs because in this environment there is same-sex or different-sex competition in getting food, so that the increase in body size makes fish more agile in moving considering the limited amount of food available. In addition, a larger size is also useful for surviving other predators such as defending or avoiding predators.

The presence of a fish population in a place and the distribution of these fish species is always related to the problem of their habitat and resources. The proportion of food availability is sufficient for the population units that live in it. So that the growth is relatively good. This means that the increase in body shape is followed by the addition of other morphometric characters, especially in total length and body weight. Then based on the DHGLM-SVM technique obtained 0.997319 by sampling method: 10-fold cross validation best parameter epsilon 0.7 cost 4 with best performance: 0.2327401. Then obtained the predictive value of fish sex with 0.997319 and Figure 5 represents the area under the curve was 0.8967.

B. EMPLOYING IN UCI MACHINE LEARNING REPOSITORY DATASET

For the next case study, we use three datasets from UCI machine learning repository. First, the bank marketing dataset published in 2012 with 45.211 instances and 17 features. Second, the car evaluation database in 1997 with 1728 instances and six features. Third, human activity recognition using smartphones Dataset. For more detail, see [4]. Table 3 and Table 4 uses seven predictors and two classes (No and Yes) with 36170 samples. Then we get the best model is SVM-DHGLM with accuracy 0.927 and 16 features. Therefore, we run the combination of random forest (RF) and SVM-DHGLM, the value of Precision was 0.937, and Recall was 0.982 and 7 best features.

Long story short, the same instrument by using the Car Evaluation Dataset. In previous study [4] using the resampling cross-validation setting of 10-fold and the tuning
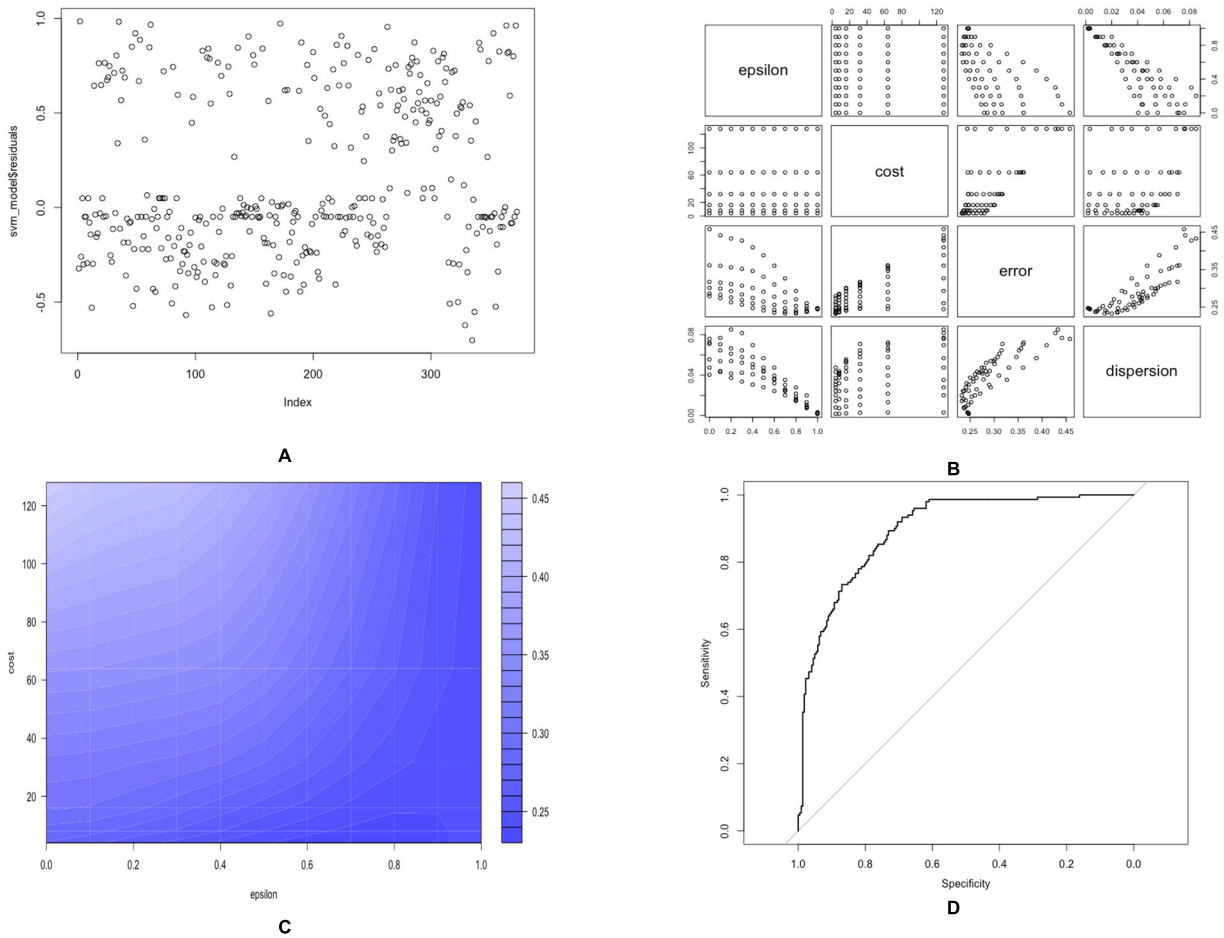
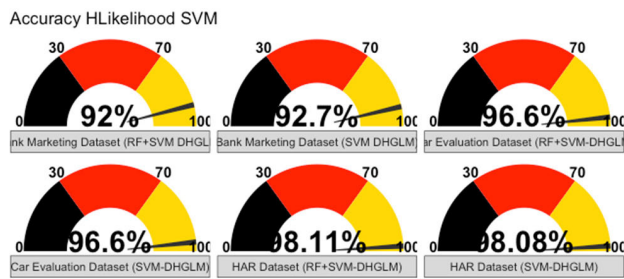**FIGURE 5.** Residual model (A), dispersion (B), performance cost and epsilon (C), and area under the curve (D).

FIGURE 6. Accuracy *H*-likelihood SVM.

TABLE 3. Classification accuracy of different classifiers with bank marketing dataset.

Method	Accuracy	Features
SVM-DHGLM (*ours)	**0.927**	**16**
SVM	0.902	16
Boruta + SVM	0.9024	7
RFE + SVM	0.9024	7
RF + SVM	0.89	7
LDA	0.8993	16
Boruta + LDA	0.9002	7
RFE + LDA	0.9002	7
RF + LDA	0.9	7
KNN	0.8877	16
Boruta + KNN	0.8874	7
RFE + KNN	0.8875	7
RF + KNN	0.886	7
RF	0.9088	16
Boruta + RF	0.9061	7
RFE + RF	0.9079	7
RF + RF	0.9099	7

parameter "sigma" was held constantly at a value of 0.07348688, C = 0.5 reach the accuracy 0.8346161, and kappa 0.6319634. For this cases, we run the different step by employing double HGLM, the construction step can be seen in the appendix. Table 5 explains the best model classifier is SVM-DHGLM with 6 features of importance and accuracy was 0.966. In the previous research [4], the best model was RF+RF with an accuracy of 0.9336. After we added the DHGLM accuracy was 0.966. Therefore, this technique provides higher accuracy.

Table 7 and Table 8 use Human Activity Recognition Using Smartphones Dataset in 2012 with 10299 instances and 561 features. In previous studies [4], the best model was SVM with accuracy 0.9796. In this article, we improvised

TABLE 4. Statistics by the class of different classifiers with bank marketing dataset.

Method	Accuracy	Precision	Recall	Features
RF+SVM	**0.92**	**0.937**	**0.982**	7
DHGLM (*ours)				
RF + SVM	0.89	0.9137	0.9791	7
RF + LDA	0.9	0.9226	0.9679	7
RF + KNN	0.886	0.9080	0.9691	7
RF + RF	0.9099	0.9122	0.9810	7

TABLE 5. Classification accuracy of different classifiers with car evaluation dataset.

Method	Accuracy	Features
SVM-DHGLM (*ours)	**0.966**	6
SVM	0.933	6
Boruta + SVM	0.8866	4
RFE + SVM	0.8895	4
RF + SVM	0.8401	4
LDA	0.8808	6
Boruta + LDA	0.843	4
RFE + LDA	0.8372	4
RF + LDA	0.843	4
KNN	0.8634	6
Boruta + KNN	0.8663	4
RFE + KNN	0.8808	4
RF + KNN	0.8227	4
RF	0.9331	6
Boruta + RF	0.8792	4
RFE + RF	0.8779	4
RF + RF	0.9336	4

TABLE 6. Classification accuracy of different classifiers with car evaluation dataset.

	Model Measurement			
	RF+SVM Accuracy: 0.8401			
Evaluation	Class: acc	Class: good	Class: unacc	Class: vgood
Precision	0.6095	NA	0.9646	0.53846
Recall	0.8421	0	0.9008	0.53846
	RF+RF Accuracy: 0.9336			
Evaluation	Class: acc	Class: good	Class: unacc	Class: vgood
Precision	0.9054	0.55	0.9832	0.75
Recall	0.8816	0.84615	0.9669	0.69231
	RF+SVM-DHGLM (*ours) Accuracy: 0.966			
Evaluation	Class: acc	Class: good	Class: unacc	Class: vgood
Precision	0.972	0.944	0.9731	0.92
Recall	0.955	0.931	0.957	0.93

TABLE 7. Classification accuracy of different classifiers with human activity recognition using smartphones dataset.

Method	Accuracy	Features
SVM-DHGLM (*ours)	0.9808	561
SVM	0.9796	561
Boruta + SVM	0.8692	6
RFE + SVM	0.8331	6
RF + SVM	0.8685	6
LDA	0.9823	561
Boruta + LDA	0.7786	6
RFE + LDA	0.705	6
RF + LDA	0.8297	6
KNN	0.9748	561
Boruta + KNN	0.864	6
RFE + KNN	0.8385	6
RF + KNN	0.904	6
RF	0.9857	561
Boruta + RF	0.8924	6
RFE + RF	0.9394	6
RF + RF	0.9326	6
RF+SVM-DHGLM (*ours)	**0.9811**	6

TABLE 8. Statistics by the class of different classifiers with human activity recognition using Smartphones dataset (6 features).

	Model Measurement	
	RF+SVM (accuracy=0.8297)	
Evaluation	Precision	Recall
Class: LAYING	1	1
Class: SITTING	0.8604	0.8872
Class: STANDING	0.891	0.865
Class: WALKING	0.7527	0.8571
Class: WALKING DOWNSTAIRS	0.8966	0.7919
Class: WALKING UPSTAIRS	0.803	0.7617
	RF+RF (accuracy=0.9326)	
Evaluation	Precision	Recall
Class: LAYING	1	1
Class: SITTING	0.924	0.9455
Class: STANDING	0.9478	0.927
Class: WALKING	0.8984	0.902
Class: WALKING DOWNSTAIRS	0.9031	0.8985
Class: WALKING UPSTAIRS	0.9019	0.9019
	RF+DHGLM SVM (*ours) (accuracy=0.9811)	
Evaluation	Precision	Recall
Class: LAYING	1	1
Class: SITTING	0.933	0.968
Class: STANDING	0.961	0.93
Class: WALKING	0.91	0.918
Class: WALKING DOWNSTAIRS	0.927	0.912
Class: WALKING UPSTAIRS	0.932	0.919

it was found that by employing DHGLM increase the accuracy which represents in Figure 6. Meanwhile, DHGLM also increase precision, recall, for feature selection and classification, respectively.

V. CONCLUSION

The classified problem is defined though in a dataset of findings consists of input, predictor, and output, containing class labels, this same classifier seeks determine the relation between the predictors and the response that will require the classification of a new observation $X \epsilon R^d$ towards one of A_z hands to form. The true class is represented by $a \epsilon A = \{1, 2, \dots, A_z\}$. Therefore, the objective of label-

with SVM-DHGLM so that the accuracy was 0.9811. At the same time, previous research conducted a combination of RF+SVM with accuracy of 0.8297, RF+RF with accuracy of 0.9326. Then, after improvising RF+DHGLM-SVM the accuracy was 0.952.

Based on the simulation using the UCI machine learning repository, namely Bank Marketing, Car Evaluation Database, Human Activity Recognition Using Smartphones,

ing is to mitigate classification errors or even the estimated consequences of misclassifications in regard for input data where some categories of mistake are more significant than some others. The cost matrix $z(u, t)$, $u, t = 1, 2, \ldots, A_z$ which describes the cost of misclassifying a representative of the class to the class. Going smoothly, Support vector machine might essentially read and understand the training set rather than learn a more general classification technique. Future research can use robust adjusted likelihood. For the likelihood ratio statistics, it is well known that, under certain regularity conditions, the classical Fisher information has the equivalent form $F(\theta) = D(\theta) = A(\theta)$ where $F(\theta)$ represents fisher information, $D(\theta)$ is the variance of the fisher score and $A(\theta)$ is expected Fisher Information. Then to get new information $A(\theta).D(\theta) = L_g\left[u(\theta X) u^T(\theta; X)\right]$ and $A(\theta) = -L_g\left[\frac{\partial^2 \log L(\theta)}{\partial \theta \partial \theta^T}\right]$. This article describes a new concept on h-likelihood for support vector machines. The classification system performs on increasing the strength of conventional likelihood function is equivalent to the SVM rule in sample space. Performance increase as well as scale parameters could be determined analytically through training results. Further research can be used and applied to image processing [25]–[48], Wi-Fi sensing [49], wide area including environment [50], [51] and disaster analysis [52]. Besides, after run UCI machine learning dataset we ensure that the comparison between the proposed SVM method and the comparative methods is fair.

APPENDIX A (ALGORITHM)

We use the following estimation algorithm for model Double Negative:

1. Choose the initial values $\omega^{(0)}$, $\psi^{(0)}$, and λ_2^0. We can set $m = 0$.
2. Do the following 4 steps bellow:
   a. Given $\psi^{(m)}$ and λ_2^m, estimate $\omega^{(m+1)}$ by maximizing h;
   b. Given $\omega^{(m+1)}$ and λ_2^m, estimate $\psi^{(m)}$ by maximizing $p\omega(m+1)(h)$;
   c. Given $\omega^{(m+1)}$ and $\psi^{(m+1)}$ by maximizing $p\omega(m+1), \psi(m+1)(h)$,
   d. Set $m = m + 1$
3. If convergences is achieved, stop; otherwise, return to step 2.

As initial values of parameter for simulation for the first simulation, we take the true values of fixed parameters and zero for random effects. As the convergence criterion, the algorithm will stop if the maximum relative change of parameters estimates less than 0.000001.

APPENDIX B (SYMBOLS & ABBREVIATIONS)

See Table 9.

APPENDIX C

Besides, Eq 17 represents the hierarchical generalized linear models defined suppose that we have response

TABLE 9. Symbols & abbreviations.

Symbol	Definition
$\mathbb{R}^n$	All Vectors With n Components, Each Of Which Is A Real Number
L_D	Regular Convex.
$L_{v,b}$	Marginal Likelihood
$p_v(h)$	Adjusted Profile Log Likelihood.
ϕ_i	The Dispersion Parameter
AICc	Akaike Information Criterion
DHGLM	Double Hierarchical Generalized Linear Models
HGLMs	Hierarchical Generalized Linear Models
KDDs	Knowledge Discovery in Databases
SVM	Support Vector Machine
$F(\theta)$	Fisher Information
a	Random Effect
b	Random Parameters That are not Observed
$f(y\|v, b; \beta, \phi)$	Function Of $f(y; \lambda)$, and Function of $f(b; \alpha)$ Each (v, b) are Given y, u and is The Conditional Distribution of b Respectively
u	Random Effect
v	Random Parameters That are not Observed
α	Corresponds to the Tuning Parameter of SVM
β	Weight Coefficient Vector
γ	Fixed Effect
$\theta(\mu_i')$	Canonical Parameter
λ	Corresponds to the Tuning Parameter of SVM and Dispersion
ξ	Slack Variable

$y = (y_1, \ldots, y_n)^t$ and unobserved random variables $u = (u_1, \ldots, u_q)^t$, having $E(y_i | u) = \mu_i'$ and $Var(y_i | u) = \phi_i V(\mu_i')$ with the conditional log-likelihood for y given as u has the generalized linear model with form.

$$l(\theta', \phi; y | u) = \sum \left\{ \left[\frac{y_i \theta(\mu_i') - b\theta(\mu_i')}{\phi_i} \right] + c(y_i, \phi_i) \right\} \tag{17}$$

where ϕ_i is the dispersion parameter and $\theta(\mu_i')$ denotes the canonical parameter which can be written with this form the Eq3. Where $g()$ is the link function, X is the $n \times p$ model matrix for fixed effects β and Z is the $n \times q$ model matrix for random effects $v = g_1(u)$ for some strictly monotonic function $g_1()$. In addition, in Eq 18 the random effects u_i are independent with dispersion parameter λ_i. In line with this, for inferences Hierarchical Generalized Linear Models (HGLMs) we proposed the h-likelihood.

$$h = l(\beta, \phi; y | v) + l(\lambda; v) \tag{18}$$

More detail, the likelihood for $y|v$ and $l(\lambda; v)$ is that for v represents $l(\beta, \phi; y | v)$. The critical issue is to select the scale of random effects in h-likelihood. Following Lee and Nelder [38] we may extend their h-log likelihood for Double Hierarchical Generalized Linear Models (DHGLMs) to the model Double Negative (DN) with NIM in Eq19 and Eq20.

$$h = \sum h_i, \tag{19}$$

$$h = \sum \log f(Y_i^0, M_i, v_i, b_i) \tag{20}$$

Therefore, let l be a likelihood with nuisance effects α. At the same time, In Eq 21 we can considered a function $p_\alpha(l)$ can be defined by:

$$p_\alpha(l) = \left[l - \frac{1}{2} \log \det \left(\frac{D(l, \alpha)}{2\pi} \right) \bigg|_{\alpha = \hat{\alpha}} \right] \quad (21)$$

For fixed effects $\boldsymbol{\beta}$ the use of $\boldsymbol{p_\beta}(\boldsymbol{L})$ is the equivalents to conditioning on $\hat{\boldsymbol{\beta}}$ to eliminate the fixed nuisance parameters $\boldsymbol{\beta}$ while random effects $\boldsymbol{v}$ the use of $\boldsymbol{p_v}(\boldsymbol{h})$ is equivalent to integrating them out by using the Laplace Approximation. Therefore, this function can be used to eliminate fixed and random effects simultaneously. We showed that $\boldsymbol{p_{v,\beta}}(\boldsymbol{h})$ in mixed linear models and represent as restricted likelihood.

CODE DATA AVAILABILITY

The analysis code datasets used in this article available from the corresponding author upon reasonable request. Also. The testing data sources come from three datasets publicly available from the UCI machine learning repository. The three datasets belong to classification data that have different total instances and features. There are Bank Marketing, Car Evaluation Database, and Human Activity Recognition Using Smartphones Dataset. The link of the websites are as follows.

1. https://archive.ics.uci.edu/ml/datasets/Bank+Marketing
2. https://archive.ics.uci.edu/ml/datasets/car+evaluation
3. https://archive.ics.uci.edu/ml/datasets/human+ activity +recognition+using+smartphones

COMPETING INTERESTS

The authors declare no competing interests.

AUTHOR CONTRIBUTIONS

Conceptualization: Rezzy Eko Caraka.
Data curation: Rezzy Eko Caraka.
Formal analysis: Rezzy Eko Caraka.
Investigation: Rezzy Eko Caraka.
Methodology: Rezzy Eko Caraka, Youngjo Lee.
Software: Rezzy Eko Caraka, Youngjo Lee.
Validation: Rezzy Eko Caraka.
Visualization: Rezzy Eko Caraka.
Writing – original draft: Rezzy Eko Caraka, Youngjo Lee.
Writing – review & editing: Rezzy Eko Caraka, Youngjo Lee, Rung Ching Chen, Toni Toharudin.

REFERENCES

[1] K. J. Cios, W. Pedrycz, R. W. Swiniarski, and L. A. Kurgan, *Data Mining: A Knowledge Discovery Approach*. New York, NY, USA: Springer, 2007, doi: 10.1007/978-0-387-36795-8.

[2] J.-A. Mueller and F. Lemke, *Self-Organising Data Mining: An Intelligent Approach to Extract Knowledge From Data*. Berlin, Germany: Herstellung, Libri Books on Demand, 1999.

[3] S. Rüping, *SVM Kernels for Time Series Analysis*. Dortmund, Germany: Univ. Dortmund, 2001, p. 8, doi: 10.17877/DE290R-15237.

[4] R.-C. Chen, C. Dewi, S.-W. Huang, and R. E. Caraka, "Selecting critical features for data classification based on machine learning methods," *J. Big Data*, vol. 7, no. 1, pp. 1–26, Dec. 2020, doi: 10.1186/s40537-020-00327-4.

[5] M.-L. Zhang and Z.-H. Zhou, "A k-nearest neighbor based algorithm for multi-label classification," in *Proc. IEEE Int. Conf. Granular Comput.*, Jul. 2005, pp. 718–721, doi: 10.1109/grc.2005.1547385.

[6] A. Pasini, "Artificial neural networks for small dataset analysis," *J. Thoracic Disease*, vol. 7, no. 5, p. 953, 2015, doi: 10.3978/j.issn.2072-1439.2015.04.61.

[7] K. Malathi and R. Kavitha, "Recognition and classification of diabetic retinopathy utilizing digital fundus image with hybrid algorithms," *Int. J. Eng. Adv. Technol.*, vol. 9, no. 1, pp. 109–122, 2019, doi: 10.35940/ijeat.A1058.109119.

[8] S. R. Gunn, "Support vector machines for classification and regression," *ISIS Tech. Rep.*, vol. 14, no. 2, pp. 67–230, 1998, doi: 10.1039/b918972f.

[9] W. T. Liao, "Clustering of time series data—A survey," *Pattern Recognit.*, vol. 38, no. 11, pp. 1857–1874, 2005, doi: 10.1016/j.patcog.2005.01.025.

[10] W. Haerdle, D. D. Prastyo, and C. M. Hafner, "Support vector machines with evolutionary feature selection for default prediction," in *Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*. Oxford, U.K.: Oxford Univ. Press, 2014.

[11] D. D. Prastyo, H. A. Khoiri, S. W. Purnami, Suhartono, and S. F. Fam, "Simulation study of feature selection on survival least square support vector machines with application to health data," in *Communications in Computer and Information Science*. Singapore: Springer, 2019, pp. 34–45, doi: 10.1007/978-981-13-3441-2_3.

[12] R. E. Caraka, R. C. Chen, T. Toharudin, M. Tahmid, B. Pardamean, and R. M. Putra, "Evaluation performance of SVR genetic algorithm and hybrid PSO in rainfall forecasting," *ICIC Express Lett. B, Appl.*, vol. 11, no. 7, pp. 631–639, 2020, doi: 10.24507/icicelb.11.07.631.

[13] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statist. Comput.*, vol. 14, no. 3, pp. 199–222, Aug. 2004, doi: 10.1023/B:STCO.0000035301.49549.88.

[14] K.-R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf, "An introduction to kernel-based learning algorithms," *IEEE Trans. Neural Netw.*, vol. 12, no. 2, pp. 181–201, Mar. 2001, doi: 10.1109/72.914517.

[15] J. Gao *et al.*, "A probabilistic framework for SVM regression and error bar estimation," *Mach. Learn.*, vol. 46, pp. 71–89, 2002, doi: 10.1023/A:1012494009640.

[16] M. Pal and G. M. Foody, "Feature selection for classification of hyperspectral data by SVM," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2297–2307, May 2010, doi: 10.1109/TGRS.2009.2039484.

[17] G. Fung and J. Stoeckel, "SVM feature selection for classification of SPECT images of Alzheimer's disease using spatial information," *Knowl. Inf. Syst.*, vol. 11, no. 2, pp. 243–258, Feb. 2007, doi: 10.1007/s10115-006-0043-5.

[18] R. Chaves, J. Ramírez, J. M. Górriz, M. López, D. Salas-Gonzalez, I. Álvarez, and F. Segovia, "SVM-based computer-aided diagnosis of the Alzheimer's disease using t-test NMSE feature selection with feature correlation weighting," *Neurosci. Lett.*, vol. 461, no. 3, pp. 293–297, Sep. 2009, doi: 10.1016/j.neulet.2009.06.052.

[19] K. Shankar, S. K. Lakshmanaprabu, D. Gupta, A. Maseleno, and V. H. C. de Albuquerque, "Optimal feature-based multi-kernel SVM approach for thyroid disease classification," *J. Supercomput.*, vol. 76, no. 2, pp. 1128–1143, Feb. 2020, doi: 10.1007/s11227-018-2469-4.

[20] M. Bourel and A. M. Segura, "Multiclass classification methods in ecology," *Ecol. Indicators*, vol. 85, pp. 1012–1021, Feb. 2018, doi: 10.1016/j.ecolind.2017.11.031.

[21] S. Villon, M. Chaumont, G. Subsol, S. Villéger, T. Claverie, and D. Mouillot, "Coral reef fish detection and recognition in underwater videos by supervised machine learning: Comparison between deep learning and HOG+SVM methods," in *Advanced Concepts for Intelligent Vision Systems* (Lecture Notes in Computer Science (LNCS)), vol. 10016. 2016, doi: 10.1007/978-3-319-48680-2_15.

[22] S. Moustakidis, G. Mallinis, N. Koutsias, J. B. Theocharis, and V. Petridis, "SVM-based fuzzy decision trees for classification of high spatial resolution remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 1, pp. 149–169, Jan. 2012, doi: 10.1109/TGRS.2011.2159726.

[23] J. Huang, J. Lu, and C. X. Ling, "Comparing naive Bayes, decision trees, and SVM with AUC and accuracy," in *Proc. 3rd IEEE Int. Conf. Data Mining*, Nov. 2003, pp. 553–556, doi: 10.1109/icdm.2003.1250975.

[24] A. Patle and D. S. Chouhan, "SVM kernel functions for classification," in *Proc. Int. Conf. Adv. Technol. Eng. (ICATE)*, Jan. 2013, pp. 1–9, doi: 10.1109/ICAdTE.2013.6524743.

[25] D. C. R. Novitasari, "Detection of COVID-19 chest X-ray using support vector machine and convolutional neural network," *Commun. Math. Biol. Neurosci.*, vol. 2020, pp. 1–19, Jul. 2020.

[26] M. Karasuyama and R. Nakano, "Optimizing SVR hyperparameters via fast cross-validation using AOSVR," in *Proc. Int. Joint Conf. Neural Netw.*, Aug. 2007, pp. 1186–1191, doi: 10.1109/IJCNN.2007.4371126.

[27] R.-C. Chen and C.-H. Hsieh, "Web page classification based on a support vector machine using a weighted vote schema," *Expert Syst. Appl.*, vol. 31, no. 2, pp. 427–435, Aug. 2006, doi: 10.1016/j.eswa.2005.09.079.

[28] R. C. Chen, K. F. Cheng, Y. H. Chen, and C. F. Hsieh, "Using rough set and support vector machine for network intrusion detection system," in *Proc. 1st Asian Conf. Intell. Inf. Database Syst.*, Apr. 2009, pp. 465–470, doi: 10.1109/ACIIDS.2009.59.

[29] H. Yasin, R. E. Caraka, Tarno, and A. Hoyyi, "Prediction of crude oil prices using support vector regression (SVR) with grid search-Cross validation algorithm," *Global J. Pure Appl. Math.*, vol. 12, no. 4, pp. 3009–3020, 2016.

[30] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, vol. 13, pp. 281–305, Feb. 2012, doi: 10.5555/2188385.2188395.

[31] I. M. Johnstone and D. M. Titterington, "Statistical challenges of high-dimensional data," *Phil. Trans. Roy. Soc. A: Math., Phys. Eng. Sci.*, vol. 367, no. 1906, pp. 4237–4253, Nov. 2009, doi: 10.1098/rsta.2009.0159.

[32] M. Noh and Y. Lee, "REML estimation for binary data in GLMMs," *J. Multivariate Anal.*, vol. 98, no. 5, pp. 896–915, May 2007, doi: 10.1016/j.jmva.2006.11.009.

[33] W. Kang, M.-S. Lee, and Y. Lee, "HGLM versus conditional estimators for the analysis of clustered binary data," *Statist. Med.*, vol. 24, no. 5, pp. 741–752, Mar. 2005, doi: 10.1002/sim.1772.

[34] R. H. Myers, D. C. Montgomery, G. G. Vining, and T. J. Robinson, *Generalized Linear Models: With Applications in Engineering and the Sciences* (Wiley Series in Probability and Statistics), 2nd ed. Hoboken, NJ, USA: Wiley, 2012.

[35] Q. V. Nong, C. T. Ng, W. Lee, and Y. Lee, "Hypothesis testing via a penalized-likelihood approach," *J. Korean Stat. Soc.*, vol. 48, no. 2, pp. 265–277, Jun. 2019, doi: 10.1016/j.jkss.2018.11.005.

[36] S. Jin, M. Noh, and Y. Lee, "H-likelihood approach to factor analysis for ordinal data," *Struct. Equation Model., A Multidisciplinary J.*, vol. 25, no. 4, pp. 530–540, Jul. 2018, doi: 10.1080/10705511.2017.1403287.

[37] I. D. Ha, M. Noh, and Y. Lee, "H-likelihood approach for joint modeling of longitudinal outcomes and time-to-event data," *Biometrical J.*, vol. 59, no. 6, pp. 1122–1143, Nov. 2017, doi: 10.1002/bimj.201600243.

[38] Y. Lee and J. A. Nelder, "Hierarchical generalized linear models," *J. Roy. Stat. Soc., B (Methodol.)*, vol. 58, no. 4, pp. 619–656, Nov. 1996, doi: 10.1111/j.2517-6161.1996.tb02105.x.

[39] S. Jin and Y. Lee, "A review of h-likelihood and hierarchical generalized linear model," *Wiley Interdiscipl. Rev., Comput. Statist.*, no. e1527, doi: 10.1002/wics.1527.

[40] Z. Shun and P. McCullagh, "Laplace approximation of high dimensional integrals," *J. Roy. Stat. Soc., B (Methodol.)*, vol. 57, no. 4, pp. 749–760, Nov. 1995, doi: 10.1111/j.2517-6161.1995.tb02060.x.

[41] Rice University. (2008). *Laplace Approximation*. [Online]. Available: https://www.ece.rice.edu/ vc3/elec633/graphical_models_notes_091108.pdf

[42] R. E. Caraka, "Using multivariate generalized linear latent variable models to measure the difference in event count for stranded marine animals," *Global J. Environ. Sci. Manage.*, vol. 7, no. 1, pp. 117–130, 2021, doi: 10.22034/gjesm.2021.01.09.

[43] Y. Lee and I. D. Ha, "Orthodox BLUP versus h-likelihood methods for inferences about random effects in tweedie mixed models," *Statist. Comput.*, vol. 20, no. 3, pp. 295–303, Jul. 2010, doi: 10.1007/s11222-009-9122-2.

[44] Y. Lee and J. A. Nelder, "Likelihood inference for models with unobservables: Another view," *Stat. Sci.*, vol. 24, no. 3, pp. 255–269, Aug. 2009, doi: 10.1214/09-STS277.

[45] Y. Lee, J. A. Nelder, and M. Noh, "H-likelihood: Problems and solutions," *Statist. Comput.*, vol. 17, no. 1, pp. 49–55, Apr. 2007, doi: 10.1007/s11222-006-9006-7.

[46] Y. Lee, L. Rönnegård, and M. Noh, *Data Analysis Using Hierarchical Generalized Linear Models With R*. 2017.

[47] T. Hastie, *The SVMpath Package*. Boca Raton, FL, USA: CRC Press, 2006.

[48] S.-K. Tai, C. Dewi, R.-C. Chen, Y.-T. Liu, X. Jiang, and H. Yu, "Deep learning for traffic sign recognition based on spatial pyramid pooling with scale analysis," *Appl. Sci.*, vol. 10, no. 19, p. 6997, Oct. 2020.

[49] A. Khalili, A.-H. Soliman, M. Asaduzzaman, and A. Griffiths, "Wi-Fi sensing: Applications and challenges," *J. Eng.*, vol. 2020, no. 3, pp. 87–97, Mar. 2020, doi: 10.1049/joe.2019.0790.

[50] N. Masseran and M. A. M. Safari, "Modeling the transition behaviors of PM_{10} pollution index," *Environ. Monit. Assessment*, vol. 192, p. 441, Jun. 2020.

[51] R. E. Caraka, R. C. Chen, T. Toharudin, B. Pardamean, H. Yasin, and S. H. Wu, "Prediction of status particulate matter 2.5 using state Markov chain stochastic process and HYBRID VAR-NN-PSO," *IEEE Access*, vol. 7, pp. 161654–161665, 2019, doi: 10.1109/ACCESS.2019.2950439.

[52] M. Asaduzzaman and A. H. M. M. Latif, "A parametric Markov renewal model for predicting tropical cyclones in bangladesh," *Natural Hazards*, vol. 73, no. 2, pp. 597–612, Sep. 2014, doi: 10.1007/s11069-014-1101-z.

REZZY EKO CARAKA received the bachelor's degree from Universitas Diponegoro, the master's degree from The National University of Malaysia, and the Ph.D. degree from the Chaoyang University of Technology. In September 2020, he was the Postdoctoral Researcher with the Research for Basic Sciences, Department of Statistics, Seoul National University, South Korea. He has authored more than 30 peer-reviewed articles and held two copyrights. At the same time, he has already published more than ten statistics books, including *Geographically Weighted Regression*, *Spatial Data Panel*, *Introduction of GLM by using R*, and *Advanced Bioinformatics by using R*. His research interests include, but are not limited to, statistics geospatial, bioinformatics, statistics ecological biodiversity, and H-likelihood. He passed technical training in intermediate-level data science by the NVIDIA Deep Learning Institute and Bioinformatics BioRealm, USA. He received numerous awards, including the Best Paper of FREKS 2019 awarded by the Ministry of Finance Indonesia and the Financial Services Authority. During his doctoral study, he received the excellent research awards from the President of CYUT, in 2019 and 2020, due to his outstanding performance in Ph.D. research and passed the viva exam the end of fourth semester with predicate high distinction (HD).

YOUNGJO LEE is currently a Full Professor with the Department of Statistics, College of Natural Sciences, Seoul National University (SNU). He has already published five distinct books in generalized linear models with random effects and GLM likelihood. His research interests include generalized linear model, hierarchical generalized linear model, random effects, data science, and statistical software development.

RUNG-CHING CHEN (Member, IEEE) received the B.S. degree from the Department of Electrical Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan, in 1987, the M.S. degree from the Institute of Computer Engineers, National Taiwan University of Science and Technology, in 1990, and the Ph.D. degree from the Department of Applied Mathematics in Computer Science, National Chung Hsing University, in 1998. He is currently a Distinguished Professor with the Department of Information Management, Chaoyang University of Technology, Taichung City, Taiwan. His research interests include network technology, pattern recognition, and knowledge engineering, the Internet of Things (IoT) and data analysis, and applications of artificial intelligence.

TONI TOHARUDIN received the M.Sc. degree from the Katholieke Universiteit Leuven, in 2005, and the Ph.D. degree in spatial sciences from the University of Groningen, in 2010. He is currently an Associate Professor with the Department of Statistics, Universitas Padjadjaran. He acted as the Head of the Research Group in Time Series and Regression. His current research interest includes statistics.

● ● ●