mathematics

Article

Predicting the Performance of Ensemble Classification Using Conditional Joint Probability

Iqbal Murtza [1,2], Jin-Young Kim [3,*] and Muhammad Adnan [4]

[1] Education and Research Center for IoT Convergence Intelligent City Safety Platform, Chonnam National University, Gwangju 61186, Republic of Korea; iqbal.murtza@au.edu.pk
[2] Department of Creative Technologies, Faculty of Computing & AI, Air University, Islamabad 44230, Pakistan
[3] Department of Intelligent Electronics and Computer Engineering, Chonnam National University, Gwangju 61186, Republic of Korea
[4] Department of Technology and Safety, UiT the Arctic University of Norway, 9019 Tromsø, Norway; muhammad.adnan@uit.no
* Correspondence: beyondi@jnu.ac.kr

Abstract: In many machine learning applications, there are many scenarios when performance is not satisfactory by single classifiers. In this case, an ensemble classification is constructed using several weak base learners to achieve satisfactory performance. Unluckily, the construction of the ensemble classification is empirical, i.e., to try an ensemble classification and if performance is not satisfactory then discard it. In this paper, a challenging analytical problem of the estimation of ensemble classification using the prediction performance of the base learners is considered. The proposed formulation is aimed at estimating the performance of ensemble classification without physically developing it, and it is derived from the perspective of probability theory by manipulating the decision probabilities of the base learners. For this purpose, the output of a base learner (which is either true positive, true negative, false positive, or false negative) is considered as a random variable. Then, the effects of logical disjunction-based and majority voting-based decision combination strategies are analyzed from the perspective of conditional joint probability. To evaluate the forecasted performance of ensemble classifier by the proposed methodology, publicly available standard datasets have been employed. The results show the effectiveness of the derived formulations to estimate the performance of ensemble classification. In addition to this, the theoretical and experimental results show that the logical disjunction-based decision outperforms majority voting in imbalanced datasets and cost-sensitive scenarios.

Keywords: machine learning; probability theory; ensemble classification; cost-sensitive learning; binary classification

MSC: 00A71; 03B48; 68T10

check for updates

Citation: Murtza, I.; Kim, J.-Y.; Adnan, M. Predicting the Performance of Ensemble Classification Using Conditional Joint Probability. *Mathematics* **2024**, *12*, 2586. https://doi.org/10.3390/math12162586

Academic Editor: Ashkan Panahi

Received: 29 July 2024
Revised: 14 August 2024
Accepted: 16 August 2024
Published: 21 August 2024

1. Introduction

In many classification scenarios and datasets, achieving satisfactory detection performance is a critical problem [1]. In such scenarios, machine learning experts naturally move to the ensemble classification to combine multiple base classifiers (learners) to achieve satisfactory accurate decisions. In ensemble classification, majority voting has gained significant attention from the research community because of its effectiveness, simplicity, and democratic style of combining the population decisions [2]. Unluckily, the construction of ensemble classification has been empirical, i.e., first, an ensemble classification is constructed and if its performance is not satisfactory then it is discarded. Contrary to the empirical approach, there is an analytical approach which is deductive. In the analytical approach, firstly a mathematical model is formulated, and then an ensemble classifier constructed accordingly. Contrary to the objectiveless, directionless, random style, and luck-driven efforts in empirical construction, the analytical approach is purposeful, goal-oriented, and systematic. Additionally, once an analytical model is built, it is useful as it can be employed to construct future models, which is not possible in an empirical approach [3].

In many applications related to video surveillance, driving assistance, pedestrian detection, and disease diagnostics, there are some additional challenges because of their cost-sensitive nature. For instance, in surveillance of human-prohibited areas, a false negative (missing a human detection) has the worst cost compared to false positive. Similarly, cancer diagnosis is also critical in which missing a cancer tumor can result in severe damage, even death of the patient. Whereas, falsely detecting a cancer tumor in a healthy person costs only some money, which can be further identified as a normal person in further tests. Likewise, in automatic driving and pedestrian detection systems, missing a human may result in serious injury or death. Consequently, for such tasks, there is a need for a cost-sensitive detection system to meet the required objectives [4,5].

In imbalanced machine learning datasets, the number of positive and negative samples differs significantly. This results in the bias of classification decisions towards the majority class samples [6,7]. Unfortunately, many datasets from cost-sensitive applications are significantly imbalanced. Even more, the number of positive samples in these datasets is critically lesser than negative samples, which results in higher false negative rates, which has a higher penalty in cost-sensitive applications [8].

There are some cost-sensitive classification techniques available in the literature. Unluckily, these contributions mainly focus on either classification technique [9,10] or data sampling [6,11]. For example, Zadrozny et al. [12] associated some weight with class learning examples to achieve cost-sensitive learning. Likewise, Krawczyk et al. [13] uses cost-sensitive analysis for breast thermography. Whereas, Nguyen et al. [6] fixed the classifier tendency to overwhelm in favor of the majority class because of imbalanced datasets by comparing many data resampling techniques and optimizing the cost ratio. Similarly, Singh et al. [14] employed transfer learning for imbalanced breast cancer classification, but a dedicated component to handle imbalanced classification is missing in their methodology. Likewise is the work of Saleeman IV et al. [15], in which they employed Spark for multiclass imbalanced classification. In addition to these conventional learning techniques, there are few deep learning-based techniques also present in the literature. For example, Almarshdi et al. [16] proposed a hybrid deep learning solution for imbalanced classification, but unluckily, the innovation in how to tackle imbalanced data is missing.

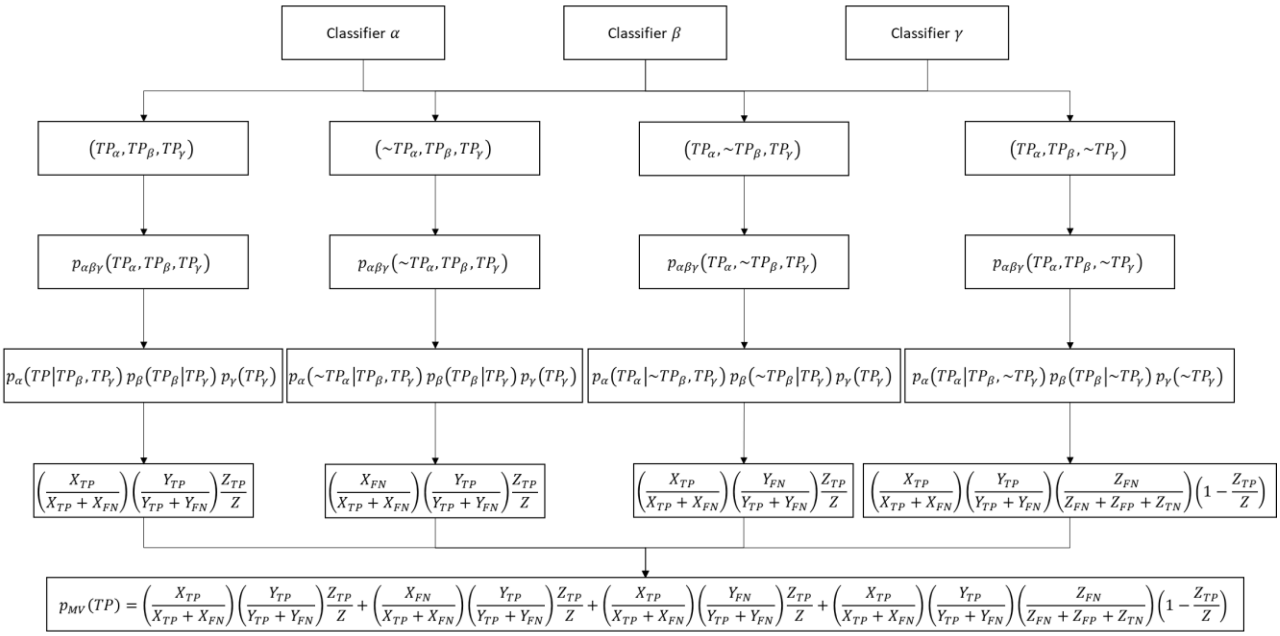
Contrary to using a single classifier [6,12,13], Liangyuan et al. [17] employed a cost-sensitive ensemble learning method using majority voting. Fan et al. [18] proposed a pruning mechanism of base classifiers, to minimize the computational cost of ensemble base cost-sensitive ensemble learning. However, these techniques do not consider the role of imbalance datasets towards the bias of the classifier [17–19].

In this line of research, some machine learning scientists focused on the role of imbalanced datasets in the designing of ensemble learning models [20]. For example, Zhang et al. [8] have proposed an ensemble method for class-imbalanced datasets by splitting the majority class dataset into various subsets and then training different base learners on minority class samples and each subset of majority class samples. Similarly, Yuan et al. [21] oversampled the dataset, used standard AdaBoost [22], and then applied genetic algorithm (GA) to optimize the weights of the base classifiers. Ali et al. [23] proposed a GentleBoost ensemble for breast cancer by oversampling the minority samples. Their work considers the probability of occurrence of each training sample to incorporate the cost effects. Hou et al. [24] employed dynamic classifier selection [25] to propose a computationally extensive dynamic ensemble classification META-DESKNN-MI. Their model uses SMOTE to fix the class imbalance in the training set. Although Xu and Chetia [26] proposed an efficient implementation of dynamic ensemble classification, unluckily, these are empirical ensemble, and additionally, ensemble selection and class imbalance are treated differently.

Unfortunately, these approaches are empirical, and thereby, they require the construction of an ensemble classifier, and if its performance is not satisfactory then discard it to try another ensemble classifier strategy. There is a significant deficiency in the literature for analytical analysis prior to the construction of ensemble classification. In this research, the problem of designing an empirical model to estimate and predict the performance of ensemble classification such as majority voting and logical disjunction is considered. For this design, the formulations are derived using the concepts of conditional joint probability. For this, the output label of a base learner has been considered as random variables with different probabilistic values for true negative (TN), false negative (FN), true positive (TP), and false positive (FP). This is an important, major, and main aspect of this research. Although, the nature of these formulations and derivations is generic, but we consider the cost-sensitive and imbalanced datasets to evaluate the forecasted performance of the ensemble classification using the derived formulations. In experiments, it is analyzed using an analytical model and experimental observations that in imbalanced datasets with cost-sensitive scenarios, logical disjunction outperforms the contemporary majority voting ensemble classification, thus providing a simple and alternative way in such scenarios.

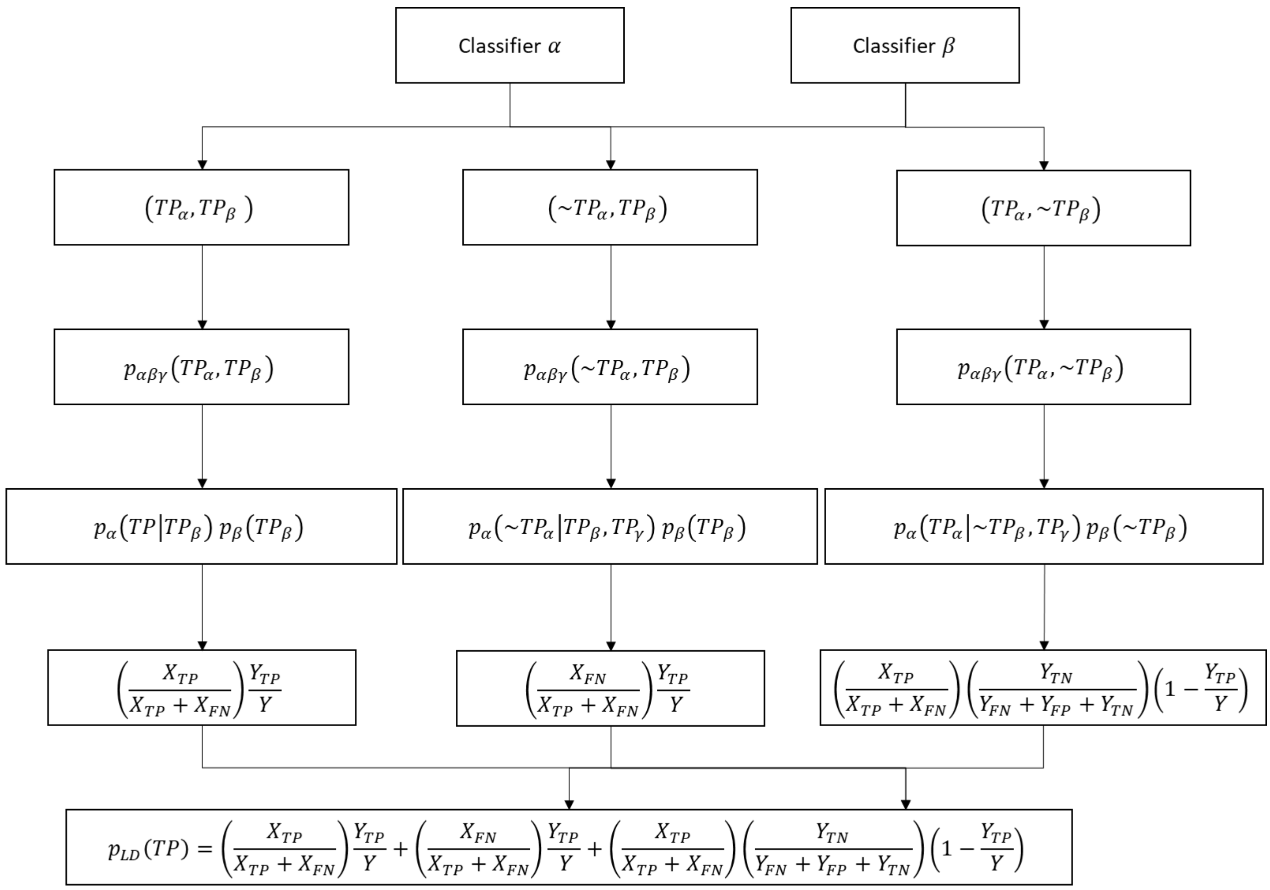
2. The Formulation to Predict the Performance of Ensemble Classification

In classification, a training dataset is used to learn feature space. After the training process is completed, a test sample is fed into the classifier and it predicts its output label. The output belongs to true positive (TP), true negative (TN), false positive (FP), or false negative (FN). It is to note that the nature of this output is random since giving a number of testing samples generates a random sequence from the set $\{TP,\ TN,\ FP, FN\}$, and thus, this set acts as sample space of this random experiment [27,28]. Using this concept, the methodology used for the derivations of the probability of true positive for ensemble classification and logical disjunction is presented in Figure 1.



(**a**)

Figure 1. *Cont.*



(**b**)

Figure 1. Analytical methodologies employed to derive the formulation of true positive probabilities for (**a**) majority voting based upon three base learners, α, β, and γ. The first layer just represents the presence of three base learners. The next layer represents the four possibilities in which majority voting gives true positive, i.e., each base learner decision is true positive $\left(TP_\alpha, TP_\beta, TP_\gamma\right)$ or any two of three base learner decisions are true positive $\left(\sim TP_\alpha, TP_\beta, TP_\gamma\right)$, $\left(TP_\alpha, \sim TP_\beta, TP_\gamma\right)$, or $\left(TP_\alpha, TP_\beta, \sim TP_\gamma\right)$. The next layer represents the probabilities of these possibilities to be computed. The next layer uses the formula of joint probability $P(x, y, z) = P(x|y, z)P(y|z)P(z)$, The next layer computes probabilities using confusion matrix (**b**) logical disjunction based upon two base learners α and β. In logical disjunction, the output decision is true positive if any base learner output is true positive. This is because, in logical disjunction, the output is positive if any base learner predicts that it is a positive sample. The description of the other layers is similar to majority voting.

2.1. Probability Perspective of Classifier Outputs

This methodology is presented for binary classification problems, wherein the output decision belongs to four categories, i.e., $TP,\ TN,\ FN$, and FP. Thus, the set $\{TP,\ TN,\ FP, FN\}$ is the sample space. Considering the output as a random variable, the probabilities (relative frequencies) of the events are in fact the classification performance measures (*TPR*, *TNR*, *FPR*, *FNR*) [29], as follows in Equation (1):

$$\begin{aligned} p(TP) = TPR = \tfrac{N_{TP}}{N} \\ p(TN) = TNR = \tfrac{N_{TN}}{N} \\ p(FP) = FPR = \tfrac{N_{FP}}{N} \\ p(FN) = FNR = \tfrac{N_{FN}}{N} \end{aligned} \quad (1)$$

2.2. Ensemble Classifiers

In machine learning, the majority voting ensemble classification technique has gained the attention of the research community because of its effectiveness and simplicity. The enhanced accuracies of ensemble classification are explained by Condorcet's jury theorem [30], which states (for binary classification):

- "If individual base classifiers have probabilities greater than 0.5 to correctly classify, then increasing the number of base classifiers, the probability of correct classification in majority voting is increased and it approaches to 1.
- If individual base classifiers have probabilities less than 0.5 to correctly classify, then increasing the number of base classifiers, the probability of correct classification in majority voting is decreased and it approaches to 0"

In addition to majority voting, this research formulates an analytical model for logical disjunction-based decision aggregation. Although, the nature of derivation is generic, considering the number of base three for majority voting (MV) and two for logical disjunction (LD) just for the sake of simplicity.

2.3. Mutual Dependency

It is to note that since the output of a classifier from the sample space $\{TP,\ TN,\ FP, FN\}$ is considered as a random variable, there is mutual dependence among the base learners. For example, if a base learner prediction is true positive, then the other base learner's prediction is either true positive or false negative. This is because, if one base learner's prediction is true positive, then it is sure that the sample is positive, and thus, the other base learner predictions are neither true negative nor false positive. Thereby, the output predictions of the base learners are not mutually independent. This important mutual dependency has to be considered while formulating conditional probability distribution for both ensemble classifications.

Consider $x,\ y,$ and z as the random variables associated with the outputs of the base classifiers α, β, and γ, respectively. Thereby, if $z = TP$, then $y|z$ (y given z) is either TP or FN. Similarly, if $z = TP$ and y is either TP or FN, then $x|y,\ z$ is also either TP or FN. These mutual dependencies are summarized in Table 1 as follows:

Table 1. Mutual dependencies of the base learner outputs. The first row represents that if prediction z of the first classifier is *TP*, then y given z $(y|z)$ can either be *TP* or *FN*. Similarly, $x|y, z$ can also be either *TP* or *FN* in this case.

Sr. #	z	$y\,|\,z$	$x\,|\,y,z$
1	TP	$\{TP, FN\}$	$\{TP, FN\}$
2	FN	$\{TP, FN\}$	$\{TP, FN\}$
3	FP	$\{FP, TN\}$	$\{FP, TN\}$
4	TN	$\{FP, TN\}$	$\{FP, TN\}$

Considering X_i, Y_i, and Z_i; $i \in \{TP,\ TN,\ FP, FN\}$ as the number of observations for the base classifiers α, β, and γ, respectively, as summarized in Table 2.

Table 2. Symbolic representation of the number of true positives, false negatives, false positives, and true negatives for the base learners.

	True Positive	False Negative	False Positive	True Negative	Total
classifiers α	X_{TP}	X_{FN}	X_{FP}	X_{TN}	$X = X_{TP} + X_{FN} + X_{FP} + X_{TN}$
classifiers β	Y_{TP}	Y_{FN}	Y_{FP}	Y_{TN}	$Y = Y_{TP} + Y_{FN} + Y_{FP} + Y_{TN}$
classifiers γ	Z_{TP}	Z_{FN}	Z_{FP}	Z_{TN}	$Z = Z_{TP} + Z_{FN} + Z_{FP} + Z_{TN}$

2.4. Formulation

In majority voting of three base learners α, β, and γ, the ensemble decision is true positive if at least two base learner decisions are true positive. Thereby, in this majority voting, true positive is when either all of the three base learner outputs are true positive or any two of three base learner outputs are true positive. Considering $p_\alpha(x)$, $p_\beta(y)$, and $p_\gamma(z)$ as the probabilities mass functions of three individual classifiers outputs $x, y, z \in \{TP,\ FN,\ FP,\ TN\}$, the probability of majority voting to give true positive $p_{MV}(TP)$ is derived using the concept joint conditional probability distribution. In this equation, $p_{\alpha\beta\gamma}(TP_\alpha, TP_\beta, TP_\gamma)$ means the joint probability of the event TP_α (base learner α gives TP), TP_β (β gives TP), and TP_γ (γ gives TP). In these derivations, the formula of joint probability for three events $P(x,y,z) = P(x|y,z)P(y|z)P(z)$ is to be kept in mind. It is to note that if the output of the base learners β and γ is true positive, then surely one thing is clear, that it is a positive sample. Thus, if a sample is positive, then base learner α has only two options as output, i.e., either to declare it as true positive or false negative. Thus, $p_\alpha(TP_\alpha | TP_\beta, TP_\gamma) = \frac{X_{TP}}{X_{TP}+X_{FN}}$ and in the similar fashion $p_\beta(TP_\beta | TP_\gamma) = \frac{Y_{TP}}{Y_{TP}+Y_{FN}}$ and $p_\gamma(TP_\gamma) = \frac{Z_{TP}}{Z}$. Using these formulations, $p_{\alpha\beta\gamma}(TP_\alpha, TP_\beta, TP_\gamma)$ is to be computed as in Equation (2).

$$
\begin{aligned}
p_{\alpha\beta\gamma}(TP_\alpha, TP_\beta, TP_\gamma) &= p_\alpha(TP_\alpha | TP_\beta, TP_\gamma)\ p_\beta(TP_\beta | TP_\gamma)\ p_\gamma(TP_\gamma) \\
p_{\alpha\beta\gamma}(TP_\alpha, TP_\beta, TP_\gamma) &= \left(\frac{X_{TP}}{X_{TP}+X_{FN}}\right)\left(\frac{Y_{TP}}{Y_{TP}+Y_{FN}}\right)\frac{Z_{TP}}{Z}
\end{aligned} \tag{2}
$$

Similarly, by computing $p_{\alpha\beta\gamma}(\sim TP_\alpha, TP_\beta, TP_\gamma)$, $p_{\alpha\beta\gamma}(TP_\alpha, \sim TP_\beta, TP_\gamma)$, and $p_{\alpha\beta\gamma}(TP_\alpha, TP_\beta, \sim TP_\gamma)$, the probability of majority voting to give true positive $p_{MV}(TP)$ is to be computed as in Equation (3). Figure 1a is additionally helpful in this derivation.

$$
\begin{aligned}
p_{MV}(TP) = \quad & p_{\alpha\beta\gamma}(TP_\alpha, TP_\beta, TP_\gamma) + p_{\alpha\beta\gamma}(\sim TP_\alpha, TP_\beta, TP_\gamma) + p_{\alpha\beta\gamma}(TP_\alpha, \sim TP_\beta, TP_\gamma) \\
& + p_{\alpha\beta\gamma}(TP_\alpha, TP_\beta, \sim TP_\gamma) \\
p_{MV}(TP) = \quad & p_\alpha(TP_\alpha | TP_\beta, TP_\gamma)\ p_\beta(TP_\beta | TP_\gamma)\ p_\gamma(TP_\gamma) \\
& + p_\alpha(\sim TP_\alpha | TP_\beta, TP_\gamma)\ p_\beta(TP_\beta | TP_\gamma)\ p_\gamma(TP_\gamma) \\
& + p_\alpha(TP_\alpha | \sim TP_\beta, TP_\gamma)\ p_\beta(\sim TP_\beta | TP_\gamma)\ p_\gamma(TP_\gamma) \\
& + p_\alpha(TP_\alpha | TP_\beta, \sim TP_\gamma)\ p_\beta(TP_\beta | \sim TP_\gamma)\ p_\gamma(\sim TP_\gamma) \\
p_{MV}(TP) = \quad & \left(\frac{X_{TP}}{X_{TP}+X_{FN}}\right)\left(\frac{Y_{TP}}{Y_{TP}+Y_{FN}}\right)\frac{Z_{TP}}{Z} + \left(\frac{X_{FN}}{X_{TP}+X_{FN}}\right)\left(\frac{Y_{TP}}{Y_{TP}+Y_{FN}}\right)\frac{Z_{TP}}{Z} \\
& + \left(\frac{X_{TP}}{X_{TP}+X_{FN}}\right)\left(\frac{Y_{FN}}{Y_{TP}+Y_{FN}}\right)\frac{Z_{TP}}{Z} \\
& + \left(\frac{X_{TP}}{X_{TP}+X_{FN}}\right)\left(\frac{Y_{TP}}{Y_{TP}+Y_{FN}}\right)\left(\frac{Z_{FN}}{Z_{FN}+Z_{FP}+Z_{TN}}\right)\left(1 - \frac{Z_{TP}}{Z}\right)
\end{aligned} \tag{3}
$$

In logical disjunction of two base learners α and β, the ensemble decision is true positive if at least one base learner decision is true positive. Thereby, in this logical disjunction, true positive is when either both base learner outputs are true positive or any base learner output is true positive. Thus, the probability of logical disjunction to give true positive $p_{LD}(TP)$ is to be computed as in Equation (4).

$$
\begin{aligned}
p_{LD}(TP) = \quad & p_{\alpha\beta}(TP_\alpha, TP_\beta) + p_{\alpha\beta}(\sim TP_\alpha, TP_\beta) + p_{\alpha\beta}(TP_\alpha, \sim TP_\beta) \\
p_{LD}(TP) = \quad & p_\alpha(TP_\alpha | TP_\beta)\ p_\beta(TP_\beta) + p_\alpha(\sim TP_\alpha | TP_\beta)\ p_\beta(TP_\beta) \\
& + p_\alpha(TP_\alpha | \sim TP_\beta)\ p_\beta(\sim TP_\beta) \\
p_{LD}(TP) = \quad & \left(\frac{X_{TP}}{X_{TP}+X_{FN}}\right)\frac{Y_{TP}}{Y} + \left(\frac{X_{FN}}{X_{TP}+X_{FN}}\right)\frac{Y_{TP}}{Y} \\
& + \left(\frac{X_{TP}}{X_{TP}+X_{FN}}\right)\left(\frac{Y_{TN}}{Y_{FN}+Y_{FP}+Y_{TN}}\right)\left(1 - \frac{Y_{TP}}{Y}\right)
\end{aligned} \tag{4}
$$

In majority voting of three base learners α, β, and γ, the ensemble decision is false negative if at least two base learner decisions are false negative. Thereby, in this majority voting, false negative is when either all base learner outputs are false negative or any two base learner outputs are false negative. Thus, the probability of majority voting to give false negative $p_{MV}(FN)$ is to be computed as in Equation (5).

$$\begin{aligned}
p_{MV}(FN) = \; & p_{\alpha\beta\gamma}(FN_\alpha, FN_\beta, FN_\gamma) + p_{\alpha\beta\gamma}(\sim FN_\alpha, FN_\beta, FN_\gamma) \\
& + p_{\alpha\beta\gamma}(FN_\alpha, \sim FN_\beta, FN_\gamma) + p_{\alpha\beta\gamma}(FN_\alpha, FN_\beta, \sim FN_\gamma) \\
p_{MV}(FN) = \; & p_\alpha(FN_\alpha | FN_\beta, FN_\gamma)\; p_\beta(FN_\beta | FN_\gamma)\; p_\gamma(FN_\gamma) \\
& + p_\alpha(\sim FN_\alpha | FN_\beta, FN_\gamma)\; p_\beta(FN_\beta | FN_\gamma)\; p_\gamma(FN_\gamma) \\
& + p_\alpha(FN_\alpha | \sim FN_\beta, FN_\gamma)\; p_\beta(\sim FN_\beta | FN_\gamma)\; p_\gamma(FN_\gamma) \\
& + p_\alpha(FN_\alpha | FN_\beta, \sim FN_\gamma)\; p_\beta(FN_\beta | \sim FN_\gamma)\; p_\gamma(\sim FN_\gamma) \\
p_{MV}(FN) = \; & \left(\tfrac{X_{FN}}{X_{TP}+X_{FN}}\right)\left(\tfrac{Y_{FN}}{Y_{TP}+Y_{FN}}\right)\tfrac{Z_{FN}}{Z} + \left(\tfrac{X_{TP}}{X_{TP}+X_{FN}}\right)\left(\tfrac{Y_{FN}}{Y_{TP}+Y_{FN}}\right)\tfrac{Z_{FN}}{Z} \\
& + \left(\tfrac{X_{FN}}{X_{TP}+X_{FN}}\right)\left(\tfrac{Y_{TP}}{Y_{TP}+Y_{FN}}\right)\tfrac{Z_{FN}}{Z} \\
& + \left(\tfrac{X_{FN}}{X_{TP}+X_{FN}}\right)\left(\tfrac{Y_{FN}}{Y_{TP}+Y_{FN}}\right)\left(\tfrac{Z_{TP}}{Z_{TP}+Z_{FP}+Z_{FN}}\right)\left(1 - \tfrac{Z_{FN}}{Z}\right)
\end{aligned} \tag{5}$$

In logical disjunction of two base learners α and β, the ensemble decision is false negative if both base learner decisions are false negative. Thus, the probability of logical disjunction to give false negative $p_{LD}(FN)$ is to be computed as in Equation (6).

$$\begin{aligned}
p_{LD}(FN) = & p_{\alpha\beta}(FN_\alpha, FN_\beta) = p_\alpha(FN_\alpha | FN_\beta)\; p_\beta(FN_\beta) \\
p_{LD}(FN) = & \left(\frac{X_{FN}}{X_{TP} + X_{FN}}\right)\frac{Y_{FN}}{Y}
\end{aligned} \tag{6}$$

In majority voting of three base learners α, β, and γ, the ensemble decision is false positive if at least two base learner decisions are false positive. Thereby, in this majority voting, false positive is when either all base learner outputs are false positive or any two of three base learner outputs are false positive. Thus, the probability of majority voting to give false positive $p_{MV}(FP)$ is to be computed as in Equation (7).

$$\begin{aligned}
p_{MV}(FP) = \; & p_{\alpha\beta\gamma}(FP_\alpha, FP_\beta, FP_\gamma) + p_{\alpha\beta\gamma}(\sim FP_\alpha, FP_\beta, FP_\gamma) \\
& + p_{\alpha\beta\gamma}(FP_\alpha, \sim FP_\beta, FP_\gamma) + p_{\alpha\beta\gamma}(FP_\alpha, FP_\beta, \sim FP_\gamma) \\
p_{MV}(FP) = \; & p_\alpha(FP_\alpha | FP_\beta, FP_\gamma)\; p_\beta(FP_\beta | FP_\gamma)\; p_\gamma(FP_\gamma) \\
& + p_\alpha(\sim FP_\alpha | FP_\beta, FP_\gamma)\; p_\beta(FP_\beta | FP_\gamma)\; p_\gamma(FP_\gamma) \\
& + p_\alpha(FP_\alpha | \sim FP_\beta, FP_\gamma)\; p_\beta(\sim FP_\beta | FP_\gamma)\; p_\gamma(FP_\gamma) \\
& + p_\alpha(FP_\alpha | FP_\beta, \sim FP_\gamma)\; p_\beta(FP_\beta | \sim FP_\gamma)\; p_\gamma(\sim FP_\gamma) \\
p_{MV}(FP) = \; & \left(\tfrac{X_{FP}}{X_{FP}+X_{TN}}\right)\left(\tfrac{Y_{FP}}{Y_{FP}+Y_{TN}}\right)\tfrac{Z_{FP}}{Z} + \left(\tfrac{X_{TN}}{X_{FP}+X_{TN}}\right)\left(\tfrac{Y_{FP}}{Y_{FP}+Y_{TN}}\right)\tfrac{Z_{FP}}{Z} \\
& + \left(\tfrac{X_{TP}}{X_{FP}+X_{TN}}\right)\left(\tfrac{Y_{TN}}{Y_{FP}+Y_{TN}}\right)\tfrac{Z_{FP}}{Z} \\
& + \left(\tfrac{X_{TP}}{X_{FP}+X_{TN}}\right)\left(\tfrac{Y_{TP}}{Y_{FP}+Y_{TN}}\right)\left(\tfrac{Z_{TN}}{Z_{TP}+Z_{FN}+Z_{FP}}\right)\left(1 - \tfrac{Z_{FP}}{Z}\right)
\end{aligned} \tag{7}$$

In logical disjunction of two base learners α and β, the ensemble decision is false positive if any base learner decision is false positive. Thereby, in this logical disjunction, false positive is when either both base learner outputs are false positive or any base learner output is false positive. Thus, the probability of logical disjunction to give false positive $p_{LD}(FP)$ is to be computed as in Equation (8).

$$\begin{aligned} p_{LD}(FP) = &\ p_{\alpha\beta}(FP_\alpha, FP_\beta) + p_{\alpha\beta}(\sim FP_\alpha, FP_\beta) + p_{\alpha\beta}(FP_\alpha, \sim FP_\beta) \\ p_{LD}(FP) = &\ p_\alpha(FP_\alpha, FP_\beta)\ p_\beta(FP_\beta) + p_\alpha(\sim FP_\alpha | FP_\beta)\ p_\beta(FP_\beta) \\ &\ + p_\alpha(FP_\alpha | \sim FP_\beta)\ p_\beta(\sim FP_\beta) \\ p_{LD}(FP) = &\ \left(\frac{X_{FP}}{X_{FP}+X_{TN}}\right)\frac{Y_{FP}}{Y} + \left(\frac{X_{TN}}{X_{FP}+X_{TN}}\right)\frac{Y_{FP}}{Y} \\ &\ + \left(\frac{X_{FP}}{X_{FP}+X_{TN}}\right)\left(\frac{Y_{TN}}{Y_{TP}+Y_{FN}+Y_{TN}}\right)\left(1-\frac{Y_{FP}}{Y}\right) \end{aligned} \tag{8}$$

In majority voting of three base learners α, β, and γ, the ensemble decision is true negative if at least two base learner decisions are true negative. Thereby, in this majority voting, true negative is when either all base learner outputs are true negative or any two base learner outputs are true negative. Thus, the probability of majority voting to give true negative $p_{MV}(TN)$ is to be computed as in Equation (9).

$$\begin{aligned} p_{MV}(TN) = &\ p_{\alpha\beta\gamma}(TN_\alpha, TN_\beta, TN_\gamma) + p_{\alpha\beta\gamma}(\sim TN_\alpha, TN_\beta, TN_\gamma) + p_{\alpha\beta\gamma}(TN_\alpha, \sim TN_\beta, TN_\gamma) \\ &\ + p_{\alpha\beta\gamma}(TN_\alpha, TN_\beta, \sim TN_\gamma) \\ p_{MV}(TN) = &\ p_\alpha(TN_\alpha | TN_\beta, TN_\gamma)\ p_\beta(TN_\beta | TN_\gamma)\ p_\gamma(TN_\gamma) + p_\alpha(\sim TN_\alpha | TN_\beta, TN_\gamma)\ p_\beta(TN_\beta | TN_\gamma)\ p_\gamma(TN_\gamma) \\ &\ + p_\alpha(TN_\alpha | \sim TN_\beta, TN_\gamma)\ p_\beta(\sim TN_\beta | TN_\gamma)\ p_\gamma(TN_\gamma) \\ &\ + p_\alpha(TN_\alpha | TN_\beta, \sim TN_\gamma)\ p_\beta(TN_\beta | \sim TN_\gamma)\ p_\gamma(\sim TN_\gamma) \\ p_{MV}(TN) = &\ \left(\frac{X_{TN}}{X_{FP}+X_{TN}}\right)\left(\frac{Y_{TN}}{Y_{FP}+Y_{TN}}\right)\frac{Z_{TN}}{Z} + \left(\frac{X_{FP}}{X_{FP}+X_{TN}}\right)\left(\frac{Y_{TN}}{Y_{FP}+Y_{TN}}\right)\frac{Z_{TN}}{Z} + \left(\frac{X_{TN}}{X_{FP}+X_{TN}}\right)\left(\frac{Y_{FP}}{Y_{FP}+Y_{TN}}\right)\frac{Z_{TN}}{Z} \\ &\ + \left(\frac{X_{TN}}{X_{FP}+X_{TN}}\right)\left(\frac{Y_{TN}}{Y_{FP}+Y_{TN}}\right)\left(\frac{Z_{FP}}{Z_{TP}+Z_{FN}+Z_{FP}}\right)\left(1-\frac{Z_{TN}}{Z}\right) \end{aligned} \tag{9}$$

In logical disjunction of two base learners α and β, the ensemble decision is true negative if both base learner decisions are true negative. Thus, the probability of logical disjunction to give true negative $p_{LD}(FN)$ is to be computed as in Equation (10).

$$\begin{aligned} p_{LD}(TN) = &\ p_{\alpha\beta}(TN_\alpha, TN_\beta) = p_\alpha(TN_\alpha | TN_\beta)\ p_\beta(TN_\beta) \\ p_{LD}(TN) = &\ \left(\frac{X_{TN}}{X_{FP}+X_{TN}}\right)\frac{Y_{TN}}{Y} \end{aligned} \tag{10}$$

3. Experiments and Results

To evaluate the proposed analytical formulations to predict the performance of ensemble classification, UCI machine learning repository has been considered. To establish another interesting fact of these proposed analytical formulations, imbalanced datasets have been considered. In addition to the significant difference between the number of samples for each class, these datasets are also cost-sensitive, i.e., there is a different cost of falsely predicting a positive (minority class) or negative (majority) sample, since negative samples are in the majority, as shown in Table 3. Thereby, the base learners have a tendency to predict more towards negative class as compared to positive class and thus $p_{FN} > p_{FP}$. In addition to this, the cost of false negative is greater than false positive $c_{FN} > c_{FP}$, where negative means being healthy and positive means being diseased. In this regard, these datasets create the intensified scenario of cost-sensitive imbalanced classification $c_{FN}p_{FN} > c_{FP}p_{FP}$. From this perspective, the proposed analytical formulations are evaluated on four different datasets, as in the following subsections.

Table 3. Distribution (number of samples) of the considered UCI datasets.

Sr. #	Dataset	+ve Samples	−ve Samples	Total
1	Breast Cancer	85	201	286
2	Wilt	74	4265	4339
3	Haberman's Survival	81	225	306
4	Thoracic Surgery	70	400	470

3.1. Breast Cancer Dataset

This dataset is generated by the institute of oncology, university medical center Ljubljana, Yugoslavia. This binary dataset is described by 9 medical attributes, and it includes 85 positive (recurrence-events) and 201 negative (no-recurrence-events) instances of cancer patients [31]. The observed confusion matrices of the individual and ensemble classifiers are shown in Table 4, where the positive class means a person has breast cancer, whereas the negative class means a person is normal. Using these confusion matrices, the observed probabilities are compared with the predicted probabilities computed from the proposed formulations and are shown in Table 5.

Table 4. Observed confusion matrices of the base learners, logical disjunction, and majority voting base ensemble classification, where the positive class means a person has breast cancer, whereas the negative class means a person is normal in Breast Cancer Dataset.

Classified as →	+ve	−ve
+ve	39	46
−ve	33	168

(a) Bayes

Classified as →	+ve	−ve
+ve	45	40
−ve	40	161

(b) Decision Stamp

Classified as →	+ve	−ve
+ve	37	48
−ve	30	171

(c) Naïve Bayes

Classified as →	+ve	-ve
+ve	54	31
−ve	57	144

(d) Logical Disjunction

Classified as →	+ve	−ve
+ve	40	45
−ve	32	169

(e) Majority Voting

Table 5. Performance (probabilities) comparison of the base learners with the predicted and observed performances of logical disjunction and majority voting for Breast Cancer Dataset.

Technique	p_{TP}	p_{FN}	p_{FP}	p_{TN}	Sum
Bayes	0.1364	0.1608	0.1154	0.5874	1
Decision Stamp	0.1573	0.1399	0.1399	0.5629	1
Naïve Bayes	0.1294	0.1678	0.1049	0.5979	1
Logical Disjunction (Predicted)	0.2215	0.0757	0.2323	0.4705	1
Logical Disjunction (Observed)	0.1888	0.1084	0.1993	0.5035	1
Majority Voting (Predicted)	0.1372	0.1600	0.0542	0.6486	1
Majority Voting (Observed)	0.1399	0.1573	0.1119	0.5909	1

3.2. Wilt Dataset

This dataset was generated from a remote sensing study for detecting diseased trees using Quickbird (a satellite) imagery. This is a highly imbalanced class containing 74 positive (diseased trees) and 4265 negative (normal) instances [32]. The observed confusion matrices of the individual and ensemble classifiers are shown in Table 6, where the positive class means a tree is diseased, whereas the negative class means the tree is normal. Using these confusion matrices, the observed probabilities are compared with the predicted probabilities computed from the proposed formulations and are shown in Table 7.

Table 6. Observed confusion matrices of the base learners, logical disjunction, and majority voting base ensemble classification, where the positive class means a tree is diseased, whereas the negative class means the tree is normal in Wilt Dataset.

Classified as →	+ve	−ve
+ve	120	67
−ve	8	305

(a) LMT

Classified as →	+ve	−ve
+ve	113	74
−ve	12	301

(b) Random Committee

Classified as →	+ve	−ve
+ve	117	70
−ve	19	294

(c) Randomizable Filter

Classified as →	+ve	−ve
+ve	133	54
−ve	17	296

(d) ical Disjunction

Classified as →	+ve	−ve
+ve	116	71
−ve	10	303

(e) Majority Voting

Table 7. Performance (probabilities) comparison of the base learners with the predicted and observed performances of logical disjunction and majority voting for Wilt Dataset.

Technique	p_{TP}	p_{FN}	p_{FP}	p_{TN}	**Sum**
LMT	0.2400	0.1340	0.0160	0.6100	1
Random Committee	0.2260	0.1480	0.0240	0.6020	1
Randomizable	0.2340	0.1400	0.0380	0.5880	1
Logical Disjunction (Predicted)	0.3210	0.0530	0.0394	0.5866	1
Logical Disjunction (Observed)	0.2660	0.108	0.0340	0.5920	1
Majority Voting (Predicted)	0.2551	0.1189	0.0030	0.6230	1
Majority Voting (Observed)	0.3320	0.1420	0.0200	0.6060	1

3.3. Haberman's Survival Dataset

This dataset is about the survival of the patients of Billings Hospital Chicago who underwent breast surgery because of cancer. This dataset is described by three features, and it includes 81 positive (the patient died within 5 years after the surgery) and 225 negative (the patient survived 5 years or longer after the surgery) instances [33]. The observed confusion matrices of the individual and ensemble classifiers are shown in Table 8, where the positive class means a patient will survive after treatment, whereas the negative class means the patient will not survive after treatment. Using these confusion matrices, the observed probabilities are compared with the predicted probabilities computed from the proposed formulations and are shown in Table 9.

Table 8. Observed confusion matrices of the base learners, logical disjunction and majority voting base ensemble classification, where the positive class means a patient will survive after treatment, whereas the negative class means the patient will not survive after treatment in Haberman's Survival Dataset.

Classified as →	+ve	−ve
+ve	29	52
−ve	25	200

(a) JRIP

Classified as →	+ve	−ve
+ve	35	46
−ve	34	191

(b) Naïve Bayes Multinominal

Classified as →	+ve	−ve
+ve	36	45
−ve	36	189

(c) Logit Boost

Classified as →	+ve	−ve
+ve	35	46
−ve	34	191

(d) Logical Disjunction

Classified as →	+ve	−ve
+ve	34	47
−ve	27	198

(e) Majority Voting

Table 9. Performance (probabilities) comparison of the base learners with the predicted and observed performances of logical disjunction and majority voting for Haberman's Survival Dataset.

Technique	p_{TP}	p_{FN}	p_{FP}	p_{TN}	**Sum**
JRip	0.0948	0.1699	0.0817	0.6536	1
Logit Boost	0.1144	0.1503	0.1111	0.6242	1
Naïve Bayes Multi-nominal	0.1177	0.1471	0.1177	0.6177	1
Logical Disjunction (Predicted)	0.1812	0.0835	0.2110	0.5243	1
Logical Disjunction (Observed)	0.1144	0.1503	0.1111	0.6242	1
Majority Voting (Predicted)	0.0975	0.1672	0.0392	0.6960	1
Majority Voting (Observed)	0.1111	0.1536	0.0882	0.6471	1

3.4. Thoracic Surgery Dataset

This dataset is about the survival of the patients of the Wroclaw Thoracic Surgery Center who underwent major lung resections because of primary lung cancer. This dataset contains 70 positive (patient died within 1 year after the surgery) and 400 negative (patients survived 1 year or longer after the surgery) instances. The observed confusion matrices of the individual and ensemble classifiers are shown in Table 10, where the positive class means a patient will survive after treatment, whereas the negative class means the patient will not survive after treatment. Using these confusion matrices, the observed probabilities are compared with the predicted probabilities computed from the proposed formulations and are shown in Table 11.

Table 10. Observed confusion matrices of the base learners, logical disjunction and majority voting base ensemble classification, where the positive class means a patient will survive after treatment, whereas the negative class means the patient will not survive after treatment in Thoracic Surgery Dataset.

Classified as →	+ve	−ve
+ve	15	55
−ve	43	357

(a) Multilayer Perceptron

Classified as →	+ve	−ve
+ve	11	59
−ve	42	358

(b) Naïve Bayes

Classified as →	+ve	−ve
+ve	10	60
−ve	47	353

(c) IBK

Classified as →	+ve	−ve
+ve	19	51
−ve	73	327

(d) Logical Disjunction

Classified as →	+ve	−ve
+ve	6	64
−ve	29	371

(e) Majority Voting

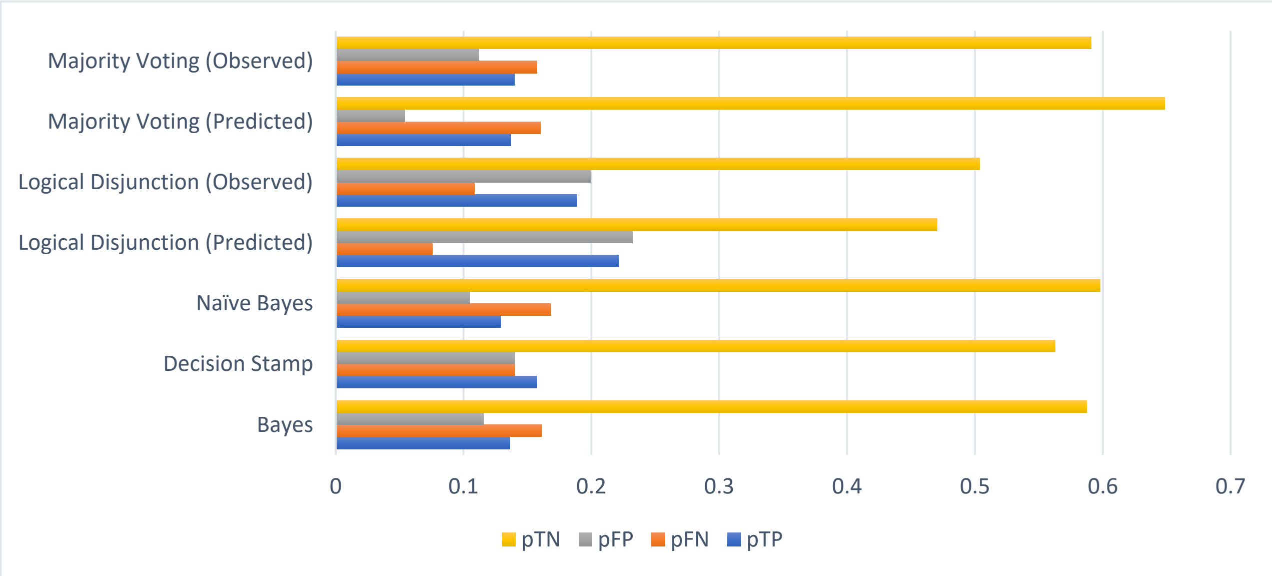
Table 11. Performance (probabilities) comparison of the base learners with the predicted and observed performances of logical disjunction and majority voting for Thoracic Surgery Dataset.

Technique	p_{TP}	p_{FN}	p_{FP}	p_{TN}	Sum
Multilayer Perceptron	0.0319	0.1170	0.0915	0.7596	1
Naïve Bayes	0.0234	0.1255	0.0894	0.7617	1
IBK	0.0213	0.1277	0.1000	0.7511	1
Logical Disjunction (Predicted)	0.0503	0.0986	0.1713	0.6798	1
Logical Disjunction (Observed)	0.0404	0.1085	0.1553	0.6957	1
Majority Voting (Predicted)	0.0115	0.1375	0.0286	0.8225	1
Majority Voting (Observed)	0.0128	0.1362	0.0617	0.7894	1

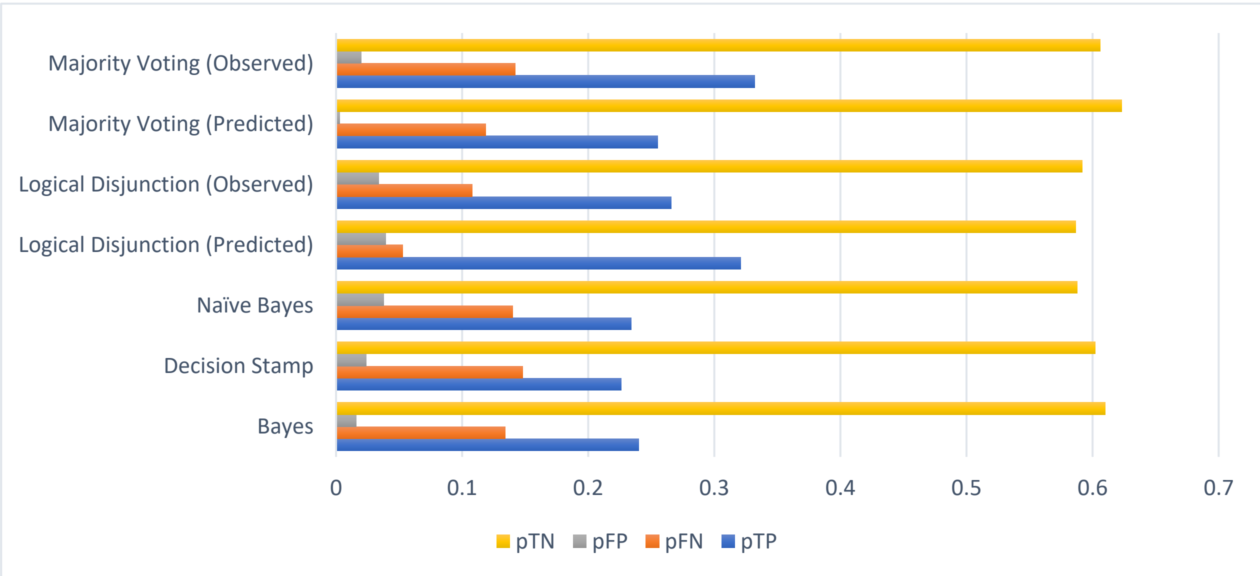
3.5. Discussion & Analysis

The experimental results in Tables 5, 7, 9 and 11 are described as graphs in Figure 2 to facilitate the comparison. From these tables and figure, it is to note that the predicted performances (p_{FN}, p_{FP}, p_{FN}, and p_{TN}) of the ensemble classifications match with the observed performance. These observations are quite encouraging and validate the effectiveness of the proposed formulations for analytical analysis prior to the actual development of ensemble classification. Thus, the proposed analytical analysis is quite helpful for deciding which base learners to be chosen and the number of base learners. The wise and early decision in this regard is useful in saving time, contrary to the empirical approach in which a model is first constructed and then continued to be discarded if not satisfactory. This is an important, major, and main aspect of the proposed formulations.

Understanding the nature of logical disjunction and majority voting base ensemble classifications, in Equations (3)–(10), it is to note that logical disjunction classifies a positive sample if any base learner classifies it as a positive sample. This is contrary to ensemble classification, which needs the majority of base learner decisions to label it as a positive sample. Thereby, it results in a decrease in the false negative rate with a tradeoff of an increase in the false positive rate, as in Tables 5–11 and Figure 2. Understanding the cost of false negatives as compared to false positives in disease diagnosis, this increase is quite useful. In these datasets, there is the scenario of $c_{FN}p_{FN} > c_{FP}p_{FP}$, and thereby, logical disjunction has been beneficial. If there is a contrary scenario of $c_{FN}p_{FN} < c_{FP}p_{FP}$, then logical conjunction is beneficial.

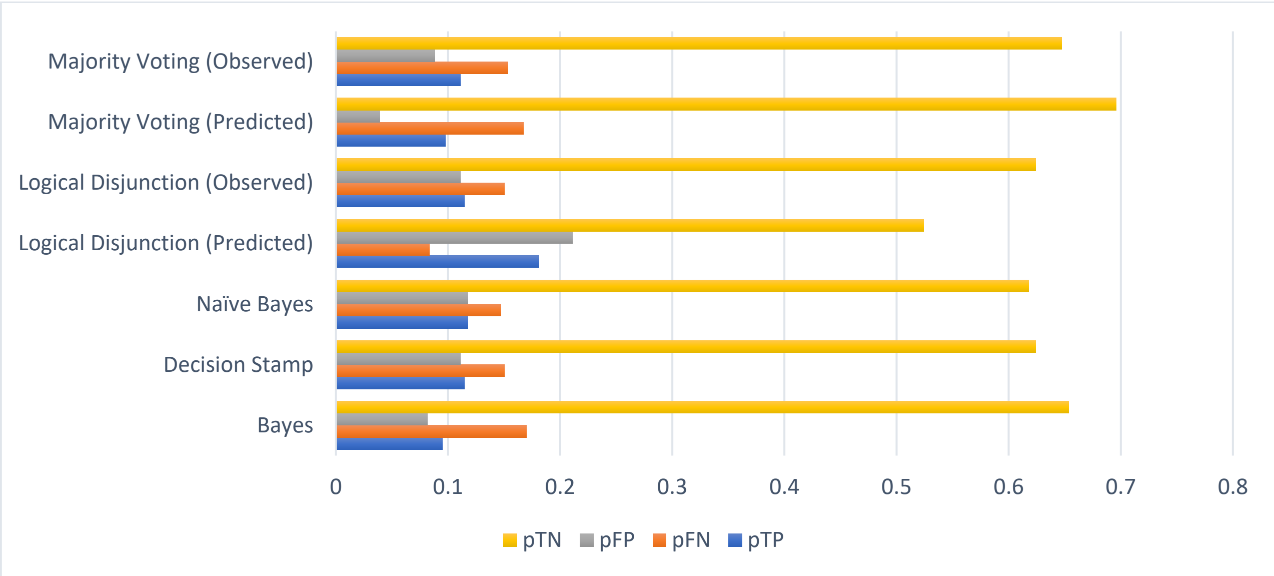


(**a**)

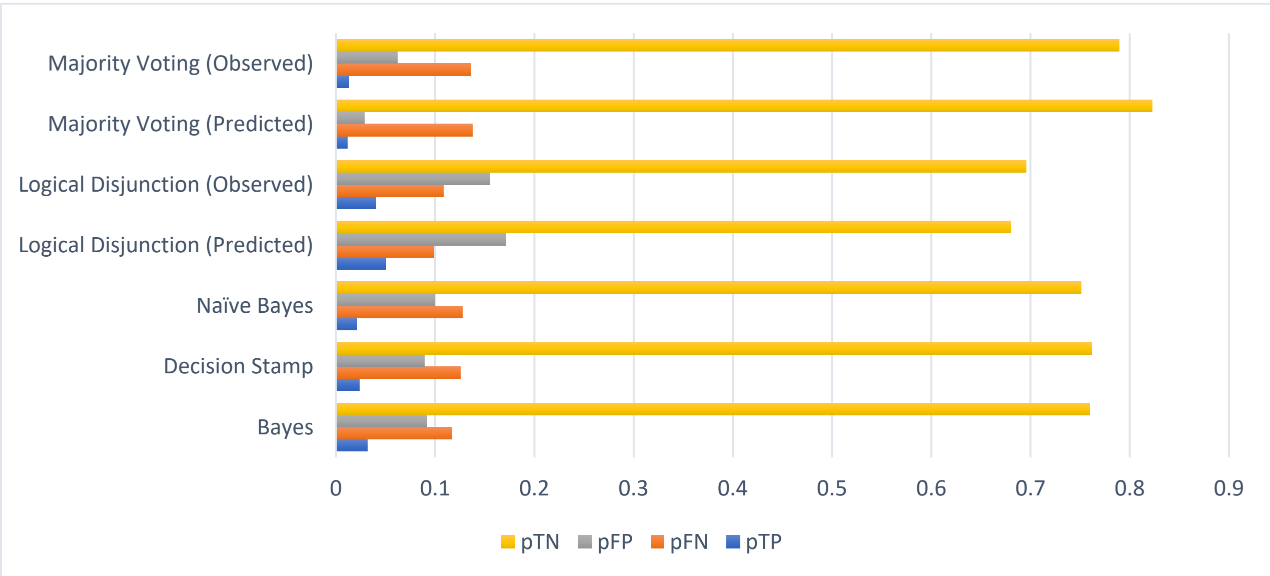


(**b**)

Figure 2. *Cont.*



(**c**)



(**d**)

Figure 2. Graphical comparison of the predicted and the observed performance of majority voting and logical disjunction-based ensemble classification techniques for the (**a**) Breast Cancer Dataset, (**b**) Wilt Dataset, (**c**) Haberman's Survival Dataset, and (**d**) Thoracic Surgery Dataset.

3.6. Conclusions

This research initiates from the consideration of true positive rate, false negative rate, false positive rate, and true negative rate as probabilities of base learners. Using this information, the concept of conditional joint probability has been applied to derive the analytical model to predict the performance of ensemble classification techniques such as majority voting and logical disjunction. The derivation of the analytical model shows that the performance of such ensemble classification can be predicted even before its actual construction using the individual performances of the base learners. The experimental observations justify the prediction of this performance. This analytical approach is useful for purposeful efforts in the construction of an appropriate ensemble classifier, contrary to the empirical approach which is a trial-based mechanism. Additionally, in the analysis and comparison of the predicted and observed performances, it is observed that for highly imbalanced datasets, the choice of logical disjunction is more appropriate as compared to the conventional majority voting for ensemble classification. Furthermore, in the case of cost-sensitive classification with highly imbalanced datasets, logical disjunction is even more appropriate as compared to majority voting. This study also shows that unwanted classification effects from highly imbalanced datasets can also be fixed using logical disjunction-based ensemble classification, contrary to the conventional under-sampling and over-sampling solutions.

Author Contributions: Conceptualization: I.M.; Methodology: I.M.; Formal Analysis: I.M. and M.A.; Investigation: I.M. and M.A.; Resources: J.-Y.K.; Writing—Original Draft Preparation: I.M.; Writing—Review & Editing: I.M., M.A. and J.-Y.K.; Visualization, I.M.; Supervision, J.-Y.K.; Project Administration: J.-Y.K.; Funding Acquisition: J.-Y.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the BK21 FOUR Program (Fostering Outstanding Universities for Research, 5199991714138) funded by the Ministry of Education (MOE, Korea) and National Research Foundation of Korea (NRF).

Data Availability Statement: The authors declared that the datasets used in this research are publicly available.

Conflicts of Interest: The authors declare that they have no conflicts of interests that could have influenced the work reported in this paper.

References

1. Leevy, J.L.; Khoshgoftaar, T.M.; Bauder, R.A.; Seliya, N. A survey on addressing high-class imbalance in big data. *J. Big Data* **2018**, *5*, 42. [CrossRef]
2. He, H.; Garcia, E.A. Learning from Imbalanced Data. *IEEE Trans. Knowl. Data Eng.* **2009**, *21*, 1263–1284. [CrossRef]
3. Flynn, B.B.; Sakakibara, S.; Schroeder, R.G.; Bates, K.A.; Flynn, E.J. Empirical research methods in operations management. *J. Oper. Manag.* **1990**, *9*, 250–284. [CrossRef]
4. Elkan, C. The foundations of cost-sensitive learning. In *International Joint Conference on Artificial Intelligence*; Lawrence Erlbaum Associates Ltd.: Mahwah, NJ, USA, 2001; Volume 17, pp. 973–978.
5. Roy, D.; Roy, A.; Roy, U. Learning from Imbalanced Data in Healthcare: State-of-the-Art and Research Challenges. In *Computational Intelligence in Healthcare Informatics*; Acharjya, D.P., Ma, K., Eds.; Springer Nature: Singapore, 2024; pp. 19–32.
6. Thai-Nghe, N.; Gantner, Z.; Schmidt-Thieme, L. Cost-sensitive learning methods for imbalanced data. In Proceedings of the 2010 International Joint Conference on Neural Networks (IJCNN), Barcelona, Spain, 18–23 July 2010; pp. 1–8.
7. El Hlouli, F.Z.; Riffi, J.; Mahraz, M.A.; Yahyaouy, A.; El Fazazy, K.; Tairi, H. Credit Card Fraud Detection: Addressing Imbalanced Datasets with a Multi-phase Approach. *SN Comput. Sci.* **2024**, *5*, 173. [CrossRef]
8. Zhang, Y.; Wang, D. A Cost-Sensitive Ensemble Method for Class-Imbalanced Datasets. *Abstr. Appl. Anal.* **2013**, *2013*, 196256.
9. Cervantes, J.; Li, X.; Yu, W. Imbalanced data classification via support vector machines and genetic algorithms. *Connect. Sci.* **2014**, *26*, 335–348. [CrossRef]
10. Wang, S.; Minku, L.L.; Chawla, N.; Yao, X. Learning from data streams and class imbalance. *Connect. Sci.* **2019**, *31*, 103–104. [CrossRef]
11. Johnson, J.M.; Khoshgoftaar, T.M. Survey on deep learning with class imbalance. *J. Big Data* **2019**, *6*, 27. [CrossRef]
12. Zadrozny, B.; Langford, J.; Abe, N. Cost-sensitive learning by cost-proportionate example weighting. In Proceedings of the Third IEEE International Conference on Data Mining, (ICDM) 2003, Melbourne, FL, USA, 19–22 November 2003; pp. 435–442.
13. Krawczyk, B.; Schaefer, G.; Wozniak, M. Breast thermogram analysis using a cost-sensitive multiple classifier system. In Proceedings of the IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), Hong Kong, China, 5–7 January 2012; pp. 507–510. [CrossRef]
14. Singh, R.; Ahmed, T.; Kumar, A.; Singh, A.K.; Pandey, A.K.; Singh, S.K. Imbalanced Breast Cancer Classification Using Transfer Learning. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2021**, *18*, 83–93. [CrossRef] [PubMed]
15. Sleeman Iv, W.C.; Krawczyk, B. Multi-class imbalanced big data classification on Spark. *Knowl.-Based Syst.* **2021**, *212*, 106598. [CrossRef]
16. Almarshdi, R.; Nassef, L.; Fadel, E.; Alowidi, N. Hybrid Deep Learning Based Attack Detection for Imbalanced Data Classification. *Intell. Autom. Soft Comput.* **2023**, *35*, 297–320.
17. Liangyuan, L.; Mei, C.; Hanhu, W.; Wei, C.; Zhiyong, G. A Cost Sensitive Ensemble Method for Medical Prediction. In Proceedings of the First International Workshop on Database Technology and Applications, Hong Kong, China, 25–26 April 2009; pp. 221–224. [CrossRef]
18. Wei, F.; Fang, C.; Haixun, W.; Philip, S.Y. Pruning and dynamic scheduling of cost-sensitive ensembles. In Proceedings of the Eighteenth National Conference on Artificial Intelligence, Edmonton, AB, Canada, 28 July–1 August 2002.

19. Chakraborty, T.; Chakraborty, A.K.; Murthy, C.A. A nonparametric ensemble binary classifier and its statistical properties. *Stat. Probab. Lett.* **2019**, *149*, 16–23. [CrossRef]
20. Depto, D.S.; Rizvee, M.M.; Rahman, A.; Zunair, H.; Rahman, M.S.; Mahdy, M.R.C. Quantifying imbalanced classification methods for leukemia detection. *Comput. Biol. Med.* **2023**, *152*, 106372. [CrossRef]
21. Bo, Y.; Xiaoli, M. Sampling + reweighting: Boosting the performance of AdaBoost on imbalanced datasets. In Proceedings of the International Joint Conference on Neural Networks (IJCNN), Brisbane, Australia, 10–15 June 2012; pp. 1–6.
22. Bartlett, P.; Traskin, M. Adaboost is consistent. In *Advances in Neural Information Processing Systems*; NeurIPS: New Orleans, LA, USA, 2006; Volume 19.
23. Ali, S.; Majid, A.; Javed, S.G.; Sattar, M. Can-CSC-GBE: Developing Cost-sensitive Classifier with Gentleboost Ensemble for breast cancer classification using protein amino acids and imbalanced data. *Comput. Biol. Med.* **2016**, *73*, 38–46. [PubMed]
24. Hou, W.-H.; Wang, X.-K.; Zhang, H.-Y.; Wang, J.-Q.; Li, L. A novel dynamic ensemble selection classifier for an imbalanced data set: An application for credit risk assessment. *Knowl.-Based Syst.* **2020**, *208*, 106462. [CrossRef]
25. Cruz, R.M.O.; Sabourin, R.; Cavalcanti, G.D.C. Dynamic classifier selection: Recent advances and perspectives. *Inf. Fusion* **2018**, *41*, 195–216. [CrossRef]
26. Xu, H.; Chetia, C. An Efficient Selective Ensemble Learning with Rejection Approach for Classification. In Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, Birmingham, UK, 21–25 October 2023; pp. 2816–2825. [CrossRef]
27. Soong, T.T. *Fundamentals of Probability and Statistics for Engineers*; Chapter 2: Basic Probability Concepts, Sec. 2.2 Sample Space and Probability Measure; John Wiley & Sons: Hoboken, NJ, USA, 2004; pp. 12–13.
28. Edition, F.; Papoulis, A.; Pillai, S.U. *Probability, Random Variables, and Stochastic Processes*; McGraw-Hill Europe: New York, NY, USA, 2002.
29. Theodoridis, S.; Koutroumbas, K. Chapter 10—Supervised Learning: The Epilogue, Sections 10.2 Error-Counting Approach and 10.3 Exploiting The Finite Size of The Data Set. In *Pattern Recognition*, 4th ed.; Theodoridis, S., Koutroumbas, K., Eds.; Academic Press: New York, NY, USA, 2009; pp. 568–573.
30. Lior, R. Ensemble-based classifiers. *Artif. Intell. Rev.* **2010**, *33*, 1–39. [CrossRef]
31. Zwitter, M.; Soklic, M. *Breat Cancer Data Set*; UCI Machine Learning Respository: Irvine, CA, USA, 1988.
32. Johnson, B. *Wilt Data Set*; UCI Machine Learning Repository: Irvine, CA, USA, 2014.
33. Lim, T.-S. *Haberman's Survival Data Set*; UCI Machine Learning Repository: Irvine, CA, USA, 1999.