New Bounds on the Accuracy of Majority Voting for Multi-Class Classification

Sina Aeeneh, Nikola Zlatanov, *Member, IEEE,* Jiangshan Yu, *Member, IEEE*

***Abstract*—Majority voting is a simple mathematical function that returns the value that appears most often in a set. As a popular decision fusion technique, the majority voting function (MVF) finds applications in resolving conflicts, where a number of independent voters report their opinions on a classification problem. Despite its importance and its various applications in ensemble learning, data crowd-sourcing, remote sensing, and data oracles for blockchains, the accuracy of the MVF for the general multi-class classification problem has remained unknown. In this paper, we derive a new upper bound on the accuracy of the MVF for the multi-class classification problem. More specifically, we show that under certain conditions, the error rate of the MVF exponentially decays toward zero as the number of voters increases. Conversely, the error rate of the MVF exponentially grows towards one if these conditions are not met.**

We first explore the problem for independent and identically distributed voters where we assume that every voter follows the same conditional probability distribution for voting for different classes, given the true classification of the data point. Next, we extend our results for the case where the voters are independent but non-identically distributed. Using the derived results, we then provide a discussion on the accuracy of the truth discovery algorithms. We show that in the best-case scenarios, truth discovery algorithms operate as an amplified MVF and thereby achieve a small error rate only when the MVF achieves a small error rate, and vice versa, achieve a large error rate when the MVF also achieves a large error rate. In the worst-case scenario, the truth discovery algorithms may achieve a higher error rate than the MVF. Finally, we confirm our theoretical results using simulations.

***Index Terms*—Majority voting, multi-class classification, crowd-sourcing, data fusion, ensemble learning, truth discovery.**

I. Introduction

MAJORITY VOTING is a simple decision rule that is used to combine multiple opinions in a classification task. The majority voting function (MVF) can be implemented in different variations including unanimity majority voting, simple majority voting, weighted majority voting, and plurality majority voting [1]. From these, plurality majority voting has been the most commonly used form of the MVF [1] and is the focus of this paper. The plurality MVF, or the MVF for short in the rest of the paper, receives votes from several voters on a specific classification problem and returns the class that has received the highest number of votes. In practice, humans use majority voting to resolve disagreements in collective decision-making processes such as elections. Computer systems also employ majority voting to improve their performance in statistical machine learning [2]–[4], data crowd-sourcing [5], remote sensing [6], [7], and, more recently, in off-chain data oracles for blockchain-based smart contracts [8]–[10]. The MVF and its variations belong to a larger set of techniques for combining different opinions known as Decision Fusion Techniques (DFTs) [11]. The voters participating in a decision fusion task together with a DFT form a decision fusion system.

Ensemble classifiers are an example of a decision fusion system. Ensemble classifiers consist of several weak classifiers and a DFT, such as a MVF, that is used to combine the outputs of the base classifiers. Ensemble classifiers are used as the state-of-the-art approach to achieve a more accurate and more generalized classification model [12]. In their seminal work in [13], the author proposed a novel technique in order to improve the performance of classifiers known as boosting. Boosting is a technique for training an ensemble of weak classifiers using different subsets of the training data set. The author in [13] showed that using a large number of weak classifiers and a proper DFT could result in an arbitrarily accurate classification model, under certain conditions. Later on, many different authors proposed similar algorithms such as bagging, AdaBoost, and stacked generalization to improve the accuracy and the generalization of the system [2]–[4]. The difference between various ensemble classification algorithms lies in the way they sample the training data used for the training of the base classifiers and in the way they combine the outputs of the base classifiers. For instance, while [2] uses the MVF for combining the outputs of the base classifiers, in [4] another classifier is trained to combine the outputs of the base classifiers and obtain the final output of the system.

The MVF, and the DFTs in general, also find applications in crowd-sourcing systems [5], where a group of independent agents try to label samples from a data set. The labeled data set is often needed for training machine learning algorithms. While individual agents participating in a crowd-sourcing task could be unreliable and noisy, the combination of their opinions using a suitable DFT may result in a data set labeled with high-accuracy [5]. Similarly, multi-sensor systems also use a DFT to combine the observations of several independent and possibly erroneous sensors in order to obtain a reliable classification result. For instance, the authors in [6] study the performance of a network of body-worn acceleration sensors in recognizing human gestures. They showed that combining the classification results of many sensors using MVF significantly improves the accuracy of the system. In

Sina Aeeneh is with the Department of Electrical and Computer Systems Engineering, Monash University, Clayton, VIC 3800, Australia (e-mail: sina.aeeneh@monash.edu).

Nikola Zlatanov is with the Department of Computer Science and Engineering, Innopolis University, Innopolis, Respublika Tatarstan, Russia, 420500.

Jiangshan Yu is with the Department of Software Systems and Cybersecurity, Monash University, Clayton, VIC 3800, Australia.

another study, the authors use DFTs to combine the outputs of different types of sensors in order to obtain a more accurate classification result in remote sensing applications [7].

More recently, the idea of employing DFTs for aggregating conflicting opinions has been studied in the blockchain literature [8]–[10]. Blockchains are decentralized systems that can record information securely across a computer network. Since blockchains are isolated systems, blockchain-specific data-feed systems have been developed to read off-chain data from sensors, web pages, and human users, and record them on the blockchains when needed. In order to increase the accuracy and reliability of their services, some data feed systems utilize multiple oracle providers for each data query task [8], [9]. Since disagreement between oracle providers may occur, a DFT is needed to aggregate the reported values and obtain a unique value for each off-chain data query. For example, ASTRAEA, a data-feed system for blockchains, employs the MVF in order to determine the final value of the off-chain data query task [8]. Since the protocol of ASTRAEA is designed for binary statements, the oracle providers need to run the slow consensus mechanism several times in order to feed a blockchain with continuous or non-binary discrete-valued off-chain data. This adds a considerable amount of delay to the system, which is not favorable [14]. Chainlink is an example of a more general decentralized data-feed system that can be implemented for both binary and non-binary-valued off-chain information [9]. While Chainlink does not employ any particular DFT, it is up to the final user to define the DFT that is suitable for their application.

The complexity of designing a robust DFT arises from the fact that in many applications the decision node performing the DFT does not have access to reliable information about the level of precision of the individual voters. Unlike the MVF which focuses on simplicity and completely disregards any potential side-information, a class of DFTs, commonly known as truth discovery algorithms, try to recursively estimate the reliability of different voters using the votes they have submitted during the previous classification tasks [15], [16]. The estimated reliability index determines the weight of each voter and is updated at the beginning of each data fusion round. Once the voters propose their values for the requested parameter, the DFT combines them in proportion to the reliability index of each voter such that the opinion of a voter that is estimated to be more reliable has a higher weight in determining the final result.

The truth discovery solutions are considered to be more accurate when compared to the MVF [15], [17]. At first glance, this seems to be correct since the truth discovery solutions exploit extra information, i.e., the past votes, to guess the accuracy of each voter. However, guessing the reliability of the voters from their past votes without having access to the past ground truths, and then, combining those past votes to generate reliability indexes, creates a positive feedback loop in the system. When compared to the MVF, the positive feedback loop of the truth discovery algorithms amplifies the accuracy in the best-case scenario only when the MVF achieves a small error rate. Nevertheless, the same positive feedback loop of the truth discovery algorithm amplifies the error rate in scenarios where the MVF achieves a high error rate. In other words, in terms of their error rate, the truth discovery algorithms are simply a MVF with an amplification factor.

In this paper, we will focus on analyzing the accuracy of the MVF as a DFT under some general assumptions. Despite its simplicity and its practicality in many applications, there is a limited theoretical study on the accuracy of the MVF in the literature, most of which assume a binary classification problem where the voters are allowed to vote either for 0 or for 1. For instance, the authors in [18] explored DFT for crowdsourcing data into blockchains for binary data queries. Their analysis showed that if individual voters have an accuracy rate that is higher than 0.5, then a majority voting-based decision maker converges to the correct value with the probability that grows towards 1 as the number of voters increases. This is in line with the results in [10], [19], [20] which have investigated the MVF and derived its error rate for the binary classification problem. All of these studies, however, lack the ability to analyze the MVF in the context of the non-binary classification problem. The multi-class classifiers are more general than the binary classifiers, making them useful for a wider range of applications. However, analyzing the accuracy of a multi-class classifier is much more difficult, since the number of votes in favor of different classes are mutually dependent random variables. To the best of our knowledge, there are only two research papers that have theoretically studied the accuracy of the MVF for the non-binary classification problem [21], [22]. The author in [21] derives an upper bound and a lower bound on the error rate of the MVF for the non-binary classification problem. However, [21] obtains bounds only for the strong MVF, a.k.a. the simple MVF, where at least 50% of the voters are needed to report a class in order for the MVF to be able to decide for that class. The strong MVF does not produce a decision if less than 50% of the voters voted for the most voted class, an event that is very frequent in multi-class classifiers. Therefore, the more commonly used plurality MVF that can achieve a much lower error rate than the strong MVF [1] is not covered in [21]. The authors in [22] propose another upper bound on the error rate of the MVF for the multi-class classification problem that is applicable to the plurality MVF. However, the upper bound in [22] is derived using the union bound technique and is therefore a loose bound in general that diverges from the correct error rate as the number of voters increases.

In this paper, we theoretically analyze the accuracy of the MVF for the multi-class classification problem by utilizing results from the *balls into bins* problem [23] and applying them to the majority voting problem. Our main contributions in this paper are listed as follows:

- We derive an upper bound on the error rate of the MVF as a DFT assuming independent and identically distributed (i.i.d.) voters. The derived upper bound is shown to be tight in the numerical examples (Theorem 1 in Section III).
- We derive the necessary and sufficient conditions under which the error rate of the MVF decays towards 0 when the number of i.i.d. voters increases (Theorem 2 and its corollaries in Section III) in which case the slope of the

decay of the error rate is exponential with the number of voters (Theorem 3 in Section III).

- We generalize the results in Theorem 1, Theorem 2, and Theorem 3 to independent non-identically distributed (i.non-i.d.) voters in Section IV. More specifically, we derive an upper bound on the error rate of the system for the general case of i.non-i.d. voters (Theorem 4 in Section IV), which is shown to be tight in the numerical examples. Moreover, we show that the error rate of the MVF for i.non-i.d. voters decays towards 0 as the number of voters increases if, and only if, certain conditions are met (Theorem 5 and its corollaries in Section IV). We also show that under those conditions the slope of the decay of the error rate is exponential with the number of voters (Theorem 6 in Section IV).
- Using the results obtained throughout Sections III and IV, we compare the MVF with a large class of DFTs, known as truth discovery algorithms, and provide insights (Section V).
- Finally we provide numerical analysis for our upper bounds on the error rate of the MVF (Section VI) and conclude the paper (Section VII).

Notation: The set $\{1,2,\ldots,N\}$ is denoted by $[N]$. The notation $[N]\setminus m$ is used to denote $\{n|\, n\in[N] \text{ and } n\neq m\}$. We use uppercase symbols, such as X, to denote random variables and lowercase symbols, such as x, for their realizations. Moreover, we use the calligraphic font to denote sets, e.g. $\mathcal{X}=\{X_1,X_2,\ldots,X_N\}$, and the bold font to denote vectors and matrices, e.g., $\mathbf{X}=(X_1,X_2,\ldots,X_N)$. We use $\mathcal{E}(.)$ to denote the probabilistic events. Given two random variables, A and B, the probability of $\mathcal{E}(A=a)$ and probability of $\mathcal{E}(A=a)$ given that $\mathcal{E}(B=b)$ has occurred is denoted by $\Pr(A=a)$ and $\Pr(A=a|B=b)$, respectively. The notation $|.|$ is used for denoting both the absolute value of variables and the number of elements in sets. Notations $\lfloor.\rfloor$ and $\lceil.\rceil$ denote the flooring and ceiling functions, respectively. Moreover, we use $\mathbf{1}(.)$ to denote the indicator function. Notation $O(.)$ is used to denote the order of a function. Finally, the modified Bessel function of the first kind of order α is denoted by $I_{|\alpha|}(.)$, where

$$I_{|\alpha|}(x)=\sum_{m=0}^{\infty}\frac{1}{m!\,(m+|\alpha|)!}\left(\frac{x}{2}\right)^{2m+|\alpha|},$$

for any real positive x and integer α.

II. System Model

Let X represent the true classification of a data point that the data fusion system is trying to classify. We model X as a random variable with probability distribution $p(x)$. Let M denote the number of voters in the data fusion system. We denote the m'th voter by O_m and its subjective opinion about the value of X by Y_m. The deviation of Y_m from X is modeled by a conditional probability distribution $p(y_m|x)$. The conditional probability distribution captures the intrinsic inaccuracy of the m'th voter resulting from imperfect data reading or imperfect processing of data as well as the potential malicious behavior of the voter. Next, each voter sends its subjective value $Y_m,\forall m\in[M]$, to a data fusion block. Then, the data fusion block uses a function, $f(.)$, defined as

$$f:(Y_1,Y_2,\ldots,Y_M)\rightarrow\hat{X}, \tag{1}$$

in order to obtain $\hat{X}$, which is the estimation of the true parameter X. Finally, the value of $\hat{X}$ is reported as the final output of the data fusion system. A sketch of the considered system is shown in Fig. 1.

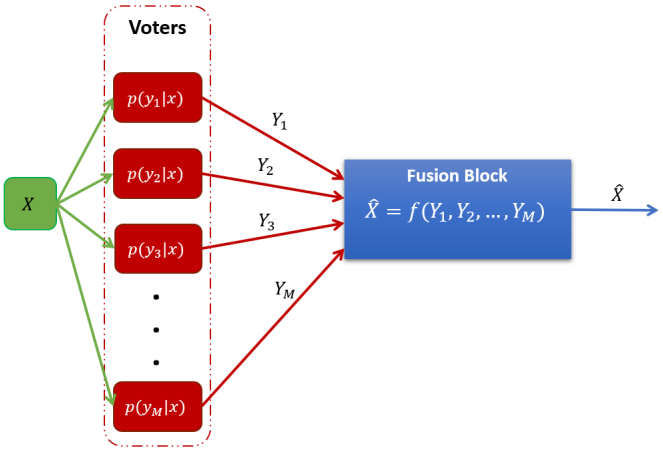

Figure 1. A sketch of the considered system model.

In the rest of this paper, we will focus on decision fusion systems that employ the MVF in the data fusion block, and thereby $f(.)$ is given by

$$f(Y_1,Y_2,\ldots,Y_M)=\underset{x_k}{\text{argmax}}\sum_{m=1}^{M}\mathbf{1}(Y_m=x_k). \tag{2}$$

Hence, the MVF first counts the number of voters that have voted for x_k, $\forall k\in[K]$, and then, at the output reports the class that has received the highest number of votes[1].

Throughout this paper we assume that X takes values from the finite set $\mathcal{X}=\{x_1,x_2,\ldots,x_K\}$. Similarly, we assume that $\hat{X}$, and Y_m, $\forall m\in[M]$, take values from the same set $\mathcal{X}$.

We use $p_{l|k}^{(m)}$ to denote the probability that a voter O_m reports $Y_m=x_l$ while $X=x_k$, i.e.,

$$p_{l|k}^{(m)}=\Pr(Y_m=x_l|X=x_k). \tag{3}$$

Our criteria for evaluating the performance of the data fusion system is the error rate defined as

$$P_e=\Pr\left(X\neq\hat{X}\right). \tag{4}$$

Hence, the optimal DFT is a DFT that minimizes P_e.

For the worst case scenario in terms of minimising P_e, we assume that X has a uniform distribution over its domain, i.e., $\Pr(X=x_k)=\frac{1}{K}$, $\forall k\in[K]$. Moreover, we assume that the voters have independent observations of X, i.e,

$$\begin{aligned}\Pr(Y_m=y_m,Y_n=y_n|X=x)&=\Pr(Y_m=y_m|X=x)\\&\times\Pr(Y_n=y_n|X=x),\ \forall m,n\in[M].\end{aligned} \tag{5}$$

[1]In the case of a tie between two or more classes, we break the tie by selecting one of the classes uniformly at random. Other rules are applicable.

This condition holds when the voters are not colluding and/or copying data from each other.

Generally, different voters may have different conditional probabilities due to the different data sources that they have access to as well as the different algorithms they may use to process data. However, in order to make the presentation easier to follow, we first investigate the case when the voters' conditional probabilities, $p\left(y_m|x\right)$'s, are i.i.d., and then we extend our results for independent but non-identically distributed (i.non-i.d.) voters.

III. Upper Bounds On The Error Rate Of The Majority Voting Function For Independent And Identically Distributed Voters

In this section, we derive an upper bound on the error rate of the MVF for i.i.d. voters. More precisely, we assume that

- the votes are independent, i.e., $\Pr\left(Y_m, Y_n|X\right) = \Pr\left(Y_m|X\right)\Pr\left(Y_n|X\right)$,
- and the votes are identically distributed, i.e., $p_{l|k}^{(m)} = p_{l|k}^{(n)},\ \forall m,n \in [M]$ and $\forall l \neq k \in [K]$.

Since we are assuming that the voters are i.i.d, the value of $p_{l|k}^{(m)}$ does not depend on the index m. Therefore, we simplify the notation and denote the conditional probability of any voter voting for $Y_m = x_l$, while $X = x_k$, by $p_{l|k}$ in the rest of this section.

A. The Upper Bound On The Error Rate

Our main result in this subsection is presented in Theorem 1, which gives an upper bound on the error rate of a DFT employing the MVF for i.i.d. voters. To this end, we first define two sets of auxiliary variables and a lemma that we need in order to prove Theorem 1.

Let $\mathbf{V} = (V_1, V_2, \ldots, V_K)$ be a random vector, where its k'th element, V_k, denotes the number of votes submitted for class x_k, and is given by

$$V_k = \sum_{m=1}^{M} \mathbf{1}\left(Y_m = x_k\right).$$

The conditional probability distribution of $\mathbf{V}$ given that $X = x_k$ is a Multinomial distribution given by

$$\Pr\left(\mathbf{V} = \mathbf{v}|X = x_k\right) = \frac{M!}{v_1!\, v_2!\, \ldots\, v_K!} \prod_{l=1}^{K} p_{l|k}^{v_l}, \tag{6}$$

where $\sum_{l=1}^{K} v_l = M$. As can be seen from (6), the random variables (RVs) $V_1, V_2, \ldots, V_K$ are mutually dependent. This makes the derivation of the error rate intractable. In order to tackle this problem, we define $\hat{\mathbf{V}} = \left(\hat{V}_1, \hat{V}_2, \ldots, \hat{V}_K\right)$ as a vector of RVs, where given $X = x_k$, the RV $\hat{V}_l \in \hat{\mathbf{V}}$ is independent of the rest of the elements in $\hat{\mathbf{V}}$ and follows a Poisson distribution with parameter $\lambda = Mp_{l|k}$, given by

$$\begin{aligned}\Pr\left(\hat{\mathbf{V}} = \hat{\mathbf{v}}|X = x_k\right) &= \prod_{l=1}^{K} \Pr\left(\hat{V}_l = \hat{v}_l|X = x_k\right) \\ &= \prod_{l=1}^{K} \frac{e^{-Mp_{l|k}}\left(Mp_{l|k}\right)^{\hat{v}_l}}{\hat{v}_l!}.\end{aligned} \tag{7}$$

We now give an important lemma that connects the probability events of $\mathbf{V}$ to the probability events of $\hat{\mathbf{V}}$.

Lemma 1: Let $\mathcal{E}\left(\mathbf{V}\right)$ be an event of a specific outcome of $\mathbf{V}$ whose probability is either monotonically increasing or monotonically decreasing with the number of voters, M. Similarly, let $\mathcal{E}\left(\hat{\mathbf{V}}\right)$ denote the same event applied on the random vector $\hat{\mathbf{V}}$. Then, the following holds

$$\Pr\left(\mathcal{E}\left(\mathbf{V}\right)\right) \leq 2\Pr\left(\mathcal{E}\left(\hat{\mathbf{V}}\right)\right). \tag{8}$$

Proof: Please refer to Appendix A. ∎

Lemma 1 plays an important role in our theoretical analysis throughout this paper by making it possible to substitute the dependent Multinomial RVs in (6) with independent Poisson RVs in (48) and obtain upper bounds on specific events. We have exploited this technique to derive an upper bound on the error rate of the MVF for i.i.d. voters, presented in the following theorem.

Theorem 1: The error rate of the majority voting DFT for i.i.d. voters is upper bounded by

$$P_e \leq 2\Bigg(1 - \frac{1}{K}\sum_{k=1}^{K}\prod_{\substack{l=1\\l\neq k}}^{K}\Bigg(1 - \sum_{\alpha=0}^{\infty} e^{-M\left(p_{k|k}+p_{l|k}\right)}\left(\frac{p_{l|k}}{p_{k|k}}\right)^{\frac{\alpha}{2}} \times I_{|\alpha|}\left(2M\sqrt{p_{l|k}p_{k|k}}\right)\Bigg)\Bigg), \tag{9}$$

where $I_{|\alpha|}\left(.\right)$ denotes the modified Bessel function of the first kind of order $|\alpha|$.

Proof: The error rate of the MVF is given by

$$\begin{aligned}P_e &= \Pr\left(\hat{X} \neq X\right) \\ &= 1 - \frac{1}{K}\sum_{k=1}^{K}\Pr\left(\hat{X} = X|X = x_k\right) \\ &\overset{(a)}{\leq} 1 - \frac{1}{K}\sum_{k=1}^{K}\Pr\left(\bigcap_{\substack{l=1\\l\neq k}}^{K} V_l < V_k \middle| X = x_k\right) \\ &\overset{(b)}{\leq} 2\left(1 - \frac{1}{K}\sum_{k=1}^{K}\Pr\left(\bigcap_{\substack{l=1\\l\neq k}}^{K} \hat{V}_l < \hat{V}_k \middle| X = x_k\right)\right) \\ &\overset{(c)}{=} 2\left(1 - \frac{1}{K}\sum_{k=1}^{K}\prod_{\substack{l=1\\l\neq k}}^{K}\Pr\left(\hat{V}_l < \hat{V}_k \middle| X = x_k\right)\right) \\ &= 2\left(1 - \frac{1}{K^{\sum_{k=1}^{K}\prod_{\substack{l=1\\l\neq k}}^{K}\left(1 - \sum_{\alpha=0}^{\infty}\Pr\left(\hat{V}_l - \hat{V}_k = \alpha \middle| X = x_k\right)\right)}}\right) \\ &\overset{(d)}{=} 2\Bigg(1 - \frac{1}{K}\sum_{k=1}^{K}\prod_{\substack{l=1\\l\neq k}}^{K}\Bigg(1 - \sum_{\alpha=0}^{\infty} e^{-M\left(p_{k|k}+p_{l|k}\right)}\left(\frac{p_{l|k}}{p_{k|k}}\right)^{\frac{\alpha}{2}} \times I_{|\alpha|}\left(2M\sqrt{p_{l|k}p_{k|k}}\right)\Bigg)\Bigg).\end{aligned} \tag{10}$$

where the right-hand side of (a) denotes the error rate assuming that MVF always breaks ties in favor of an incorrect class. Of course, this is a worst-case assumption. In practice, we can simply implement the MVF to break ties at random and achieve a lower error rate. However, in (a) we bound the error rate of the MVF with the worst-case assumption for simplicity. Next, (b) is a result of Lemma 1. Please note that the error rate, P_e, is a monotonic function of M since the voters are i.i.d. Hence, we can use Lemma 1 to bound the error rate as per (b). Next, (c) is derived from the conditional independence of the elements in $\hat{\mathbf{V}}$. Finally, (d) is obtained by substituting $\Pr\left(\hat{V}_l - \hat{V}_k = \alpha | X = x_k\right)$ with the Skellam probability mass function which is the probability mass function of the difference of two independent Poisson random variables [24]. ■

The upper bound obtained in Theorem 1 is in its most general form and may not provide intuition about its behavior. To provide intuition, we further simplify (9) to explore the behavior of the upper bound in Theorem 1 in both the asymptotic and the transient cases.

B. Asymptotic Behaviour Of The Error Rate

In the following, we first state a theorem and then exploit it to simplify the upper bound in (9) in the asymptotic scenario.

Theorem 2: Assuming that $p_{l_1|k} > p_{l_2|k}$ for given $k, l_1,$ and $l_2 \in [K]$, we can bound $\Pr\left(V_{l_1} \leq V_{l_2} | X = x_k\right)$ as

$$\Pr\left(V_{l_1} \leq V_{l_2} | X = x_k\right) \leq 2e^{-M\left(\sqrt{p_{l_1|k}} - \sqrt{p_{l_2|k}}\right)^2}. \tag{11}$$

Proof: Please refer to Appendix B. ■

Corollary 1: In the asymptotic regime, where $M \to \infty$, the error rate, conditioned on $X = x_k$, goes towards 0 if $p_{k|k} > p_{l|k}$, $\forall l \neq k$ and it goes towards 1 if $\exists\, l_0 \neq k$ such that $p_{k|k} < p_{l_0|k}$.

Proof: Assuming that $p_{l_1|k} > p_{l_2|k}$ for given $k, l_1,$ and $l_2 \in [K]$, we take the limit of both sides of the inequality in (11) as

$$\begin{aligned}\lim_{M\to\infty} \Pr\left(V_{l_1} \leq V_{l_2} | X = x_k\right) &\leq \lim_{M\to\infty} 2e^{-M\left(\sqrt{p_{l_1|k}} - \sqrt{p_{l_2|k}}\right)^2} \\ &= 0. \end{aligned} \tag{12}$$

Using (12), for the case where $p_{k|k} > p_{l|k}$, $\forall l \neq k$, we bound the error rate, conditioned on $X = x_k$, as

$$\begin{aligned}\lim_{M\to\infty} P_{e|x_k} &\overset{(a)}{\leq} \lim_{M\to\infty} \Pr\left(\bigcup_{\substack{l=1\\ l\neq k}}^{K} V_l \geq V_k | X = x_k\right) \\ &\overset{(b)}{\leq} \lim_{M\to\infty} \sum_{\substack{l=1\\ l\neq k}}^{K} \Pr\left(V_l \geq V_k | X = x_k\right) \\ &\overset{(c)}{=} 0, \end{aligned} \tag{13}$$

where the right-hand side of (a) denotes the error rate assuming that MVF always breaks ties in favor of an incorrect class, which is a worst-case assumption. Moreover, (b) is obtained using union bound, and (c) is due to (12). Conversely, if $\exists\, l_0 \neq k$ such that $p_{k|k} < p_{l_0|k}$, we have

$$\begin{aligned}\lim_{M\to\infty} P_{e|x_k} &\overset{(c)}{\geq} \lim_{M\to\infty} \Pr\left(\bigcup_{\substack{l=1\\ l\neq k}}^{K} V_l > V_k | X = x_k\right) \\ &\overset{(d}{\geq} \lim_{M\to\infty} \Pr\left(V_{l_0} > V_k | X = x_k\right) \\ &= 1 - \lim_{M\to\infty} \Pr\left(V_{l_0} \leq V_k | X = x_k\right) \\ &\overset{(e)}{=} 1, \end{aligned} \tag{14}$$

where the right-hand side of (c) denotes the error rate assuming that the MVF always breaks ties in favor of the correct class x_k. This is a best-case assumption. Therefore, the right-hand side of (c) is a lower bound on the actual error rate. Moreover, (d) results from the fact that the occurrence of the event $\mathcal{E}\left(\bigcup_{\substack{l=1\\ l\neq k}}^{K} V_l > V_k\right)$ always implies the occurrence of the event $\mathcal{E}\left(V_{l_0} > V_k\right)$. Finally, (e) is due to (12) and as a result of the fact that $p_{k|k} < p_{l_0|k}$. Combining (13) and (14) into one equation, we can describe the asymptotic behavior of the error rate, conditioned on $X = x_k$, as

$$\begin{aligned}\lim_{M\to\infty} P_{e|x_k} &\triangleq \lim_{M\to\infty} \Pr\left(\hat{X} \neq x_k | X = x_k\right) \\ &= \begin{cases} 0 & \text{if } \forall\ l \neq k,\ p_{k|k} > p_{l|k} \\ 1 & \text{if } \exists\ l \neq k,\ p_{k|k} < p_{l|k}. \end{cases} \end{aligned} \tag{15}$$

■

Corollary 2: Let $\mathcal{A}$ be the largest subset of $[K]$ such that $\forall k \in \mathcal{A}$ we have $p_{k|k} > p_{l|k}$, $\forall l \in [K]$. Then, the error rate of the MVF converges towards $1 - \frac{|\mathcal{A}|}{K}$ as M grows.

Proof: The proof is straightforward. From (15), we have

$$\begin{aligned}\lim_{M\to\infty} P_e &= \lim_{M\to\infty} \sum_{k=1}^{K} \Pr\left(X = x_k\right) P_{e|x_k} \\ &= \sum_{k\in\mathcal{A}} \frac{1}{K} \times 0 + \sum_{k\in[K]\setminus\mathcal{A}} \frac{1}{K} \\ &= 1 - \frac{|\mathcal{A}|}{K}. \end{aligned} \tag{16}$$

■

Corollary 2 is a generalization of the well-known Condorcet Jury Theorem [1] that is originally stated for the binary MVF. Due to Corollary 2, we need $p_{k|k} > p_{l|k}$, $\forall l, k \in [K]$, for a decision fusion system to work reliably, i.e.,

$$\lim_{M\to\infty} P_e = 0 \iff \forall l, k \in [K],\ \text{where } l \neq k,\ p_{k|k} > p_{l|k}. \tag{17}$$

C. Transient Behaviour Of The Error Rate

In the following, we present a theorem that exploits the inequality in (11) to simplify the upper bound obtained in Theorem 1 for the case when $p_{k|k} > p_{l|k}$, $\forall k, l \in [K]$. To this end, we first define two auxiliary variables, k^* and l^*, as

$$(k^*, l^*) = \underset{\substack{k,l\\ k\neq l}}{\operatorname{argmin}} \left(\sqrt{p_{k|k}} - \sqrt{p_{l|k}}\right). \tag{18}$$

In words, (k^*, l^*) is the pair (k,l) that minimizes the difference $\left(\sqrt{p_{k|k}} - \sqrt{p_{l|k}}\right)$, $\forall\,(k,l) \in [K]$, where $k \neq l$.

Theorem 3: If $p_{k|k} > p_{l|k}, \forall k,l \in [K]$, then the upper bound in Theorem 1 is simplified to

$$P_e \leq e^{-M\left(\left(\sqrt{p_{k^*|k^*}} - \sqrt{p_{l^*|k^*}}\right)^2 + \epsilon\right)}, \tag{19}$$

where $O(\epsilon) = \frac{\ln\left(M^{-3/4}\right)}{M}$.

Proof: Please refer to Appendix C. ∎

By taking the logarithm of both sides of (19) and dividing the result by M, we obtain

$$\frac{\ln P_e}{M} \leq -\left(\sqrt{p_{k^*|k^*}} - \sqrt{p_{l^*|k^*}}\right)^2 + \epsilon. \tag{20}$$

Therefore, assuming that the voters satisfy $p_{k|k} > p_{l|k}, \forall k,l \in [K]$, (20) shows that the slope of the decay is at most $-\left(\sqrt{p_{k^*|k^*}} - \sqrt{p_{l^*|k^*}}\right)^2$.

As we have shown above, when $p_{k|k} > p_{l|k}$, $\forall l,k \in [K]$, the MVF achieves a very small error rate assuming that the number of participating voters, M, is large enough. On the contrary, we observed that if for a given $k \in [K]$, $\exists l \neq k \in [K]$ such that $p_{k|k} < p_{l|k}$ then the MVF almost surely makes an error in detecting X when $X = x_k$.

IV. Upper Bounds On The Error Rate Of The Majority Voting Function For The Case Of Independent and Non-Identically Distributed Voters

In this section, we extend the analysis in Section III to obtain upper bounds on the error rate of the MVF for the scenario when the voters are i.non-i.d. More precisely, throughout this section, we only assume that

- the voters are independent, i.e., $\Pr(Y_m, Y_n|X) = \Pr(Y_m|X)\Pr(Y_n|X)$.

Recalling from Section II that $p_{l|k}^{(m)}$ denotes the probability that voter O_m reports $Y_m = x_l$ while $X = x_k$, we define a matrix, which we refer to it as the probability transition matrix for voter O_m and denote it by $\mathrm{p}^{(m)}$, as

$$\mathrm{p}^{(m)} = \begin{bmatrix} p_{1|1}^{(m)} & p_{1|2}^{(m)} & \cdots & p_{1|K}^{(m)} \\ p_{2|1}^{(m)} & p_{2|2}^{(m)} & \cdots & p_{2|K}^{(m)} \\ \vdots & \vdots & \ddots & \vdots \\ p_{K|1}^{(m)} & p_{K|2}^{(m)} & \cdots & p_{K|K}^{(m)} \end{bmatrix}. \tag{21}$$

We partition the voters based on their probability transition matrices. To this end, let $\{\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_T\}$ be a partitioning on the set of voters, $\mathcal{O} = \{O_1, O_2, \ldots, O_M\}$, such that every element in $\mathcal{O}$ falls in exactly one set $\mathcal{G}_t$ and for any two voters, O_m and O_n that fall in the same set, we have $\mathrm{p}^{(m)} = \mathrm{p}^{(n)}$. For simplicity of notation, the probability transition matrix of $O_m, \forall\, O_m \in \mathcal{G}_t$ is denoted by $\mathbf{q}^{(t)}$, where the (k,l) element of $\mathbf{q}^{(t)}$ is denoted by $q_{l|k}^{(t)}$ and is given by

$$q_{l|k}^{(t)} = p_{l|k}^{(m)},\ \forall\, m \in \mathcal{G}_t. \tag{22}$$

The intuition behind partitioning voters into mutually disjoint sets is to capture the differences and identities in their probability transition matrices. Here, we should emphasize that the MVF does not need to know the probability transition matrix of different voters; we use the notion of the probability transition matrix, and the partitioning introduced above, only for obtaining the upper bound on the error rate of the MVF.

Let $V_k^{(t)}$ denote the number of voters in $\mathcal{G}_t$ that have voted for class x_k. Next, we define random vectors $\mathbf{V}^{(t)}$ and $\mathbf{V}$ as

$$\mathbf{V}^{(t)} \triangleq \left(V_1^{(t)}, V_2^{(t)}, \ldots, V_K^{(t)}\right),$$
$$\mathbf{V} \triangleq \left(\mathbf{V}^{(1)}, \mathbf{V}^{(2)}, \ldots, \mathbf{V}^{(T)}\right).$$

In other words, $\mathbf{V}^{(t)}$ denotes the number of voters in $\mathcal{G}_t$ that have voted for different classes in $\mathcal{X}$ and $\mathbf{V}$ is created by concatenating $\mathbf{V}^{(t)}$'s.

Let r_t be the probability that a randomly selected voter belongs to $\mathcal{G}_t$. The conditional probability distribution of $\mathbf{V}^{(t)} = \left(V_1^{(t)}, V_2^{(t)}, \ldots, V_K^{(t)}\right)$ given $X = x_k$ is a Multinomial distribution, given by

$$\Pr\left(\mathbf{V}^{(t)} = \mathbf{v}^{(t)} \middle| X = x_k\right) = \frac{M!}{\prod_{t=1}^{T}\prod_{l=1}^{K} v_l^{(t)}} \prod_{t=1}^{T}\prod_{l=1}^{K}\left(r_t q_{l|k}^{(t)}\right)^{v_l^{(t)}}, \tag{23}$$

where $\sum_{t=1}^{T}\sum_{k=1}^{K} v_k^{(t)} = M$.

The output of the MVF is determined by finding the maximum element of $\mathbf{V} = (V_1, V_2, \ldots, V_K)$ where its k'th element V_k is given by

$$V_k \triangleq \sum_{t=1}^{T} V_k^{(t)}.$$

Please note that unlike what we had in Section III for i.i.d. voters, here the random vector $\mathbf{V}$ does not necessarily follow a Multinomial distribution, only $\mathbf{V}^{(t)}$ follows a Multinomial distribution, see (23).

A. The Upper Bound On The Error Rate

We state a lemma that is used in the proof of the upper bound on the error rate of the MVF for i.non-i.d. voters. First, we define an auxiliary variable, $\delta_{l_1,l_2|k}$, as

$$\delta_{l_1,l_2|k} \triangleq \sum_{t=1}^{T} r_t\left(q_{l_1|k}^{(t)} - q_{l_2|k}^{(t)}\right). \tag{24}$$

In words, $\delta_{l_1,l_2|k}$ measures the difference between the likeliness of a randomly selected voter voting for class x_{l_1} and the likeliness that it votes for the class x_{l_2}, given that $X = x_k$. We use $\delta_{l_2|k}$ instead of $\delta_{k,l_2|k}$ for the special case where $l_1 = k$ in order to simplify the notation.

Lemma 2: If $\hat{X} \neq X$ when $X = x_k$, then there exists at least one $l \in [K]$ and one $t \in [T]$ such that

$$v_l^{(t)} - v_k^{(t)} \geq E\left\{V_l^{(t)} - V_k^{(t)}\right\} + r_t M \delta_{l|k}, \tag{25}$$

where $v_l^{(t)}$ and $v_k^{(t)}$ are the realizations of $V_l^{(t)}$ and $V_k^{(t)}$, respectively.

Proof: Please refer to Appendix D. ∎

Lemma 2 shows that for the MVF to make an incorrect decision when $X = x_k$, the inequality in (25) must hold for at least one $t \in [T]$ and one $l \in [K]$. Next, we exploit this

$$
\begin{aligned}
P_e=&\Pr\left(\hat{X}\neq X\right)\\
=&\frac{1}{K}\sum_{k=1}^{K}\Pr\left(\hat{X}\neq X|X=x_k\right)\\
\overset{(a)}{\leq}&\frac{1}{K}\sum_{k=1}^{K}\Pr\left(\bigcup_{\substack{l=1\\l\neq k}}^{K}V_l\geq V_k\Big|X=x_k\right)\\
\overset{(b)}{=}&\frac{1}{K}\sum_{k=1}^{K}\Pr\left(\bigcup_{t=1}^{T}\bigcup_{\substack{l=1\\l\neq k}}^{K}\left\{V_l^{(t)}-V_k^{(t)}\geq E\left\{V_l^{(t)}-V_k^{(t)}\right\}+r_tM\delta_{l|k}\right\}\Big|X=x_k\right)\\
\overset{(c)}{\leq}&\frac{2}{K}\sum_{k=1}^{K}\Pr\left(\bigcup_{t=1}^{T}\bigcup_{\substack{l=1\\l\neq k}}^{K}\left\{\hat{V}_l^{(t)}-\hat{V}_k^{(t)}\geq E\left\{V_l^{(t)}-V_k^{(t)}\right\}+r_tM\delta_{l|k}\right\}\Big|X=x_k\right)\\
=&\frac{2}{K}\sum_{k=1}^{K}\left(1-\Pr\left(\bigcap_{t=1}^{T}\bigcap_{\substack{l=1\\l\neq k}}^{K}\left\{\hat{V}_l^{(t)}-\hat{V}_k^{(t)}<E\left\{V_l^{(t)}-V_k^{(t)}\right\}+r_tM\delta_{l|k}\right\}\Big|X=x_k\right)\right)\\
\overset{(d)}{=}&\frac{2}{K}\sum_{k=1}^{K}\left(1-\prod_{t=1}^{T}\prod_{\substack{l=1\\l\neq k}}^{K}\Pr\left(\hat{V}_l^{(t)}-\hat{V}_k^{(t)}<E\left\{V_l^{(t)}-V_k^{(t)}\right\}+r_tM\delta_{l|k}\Big|X=x_k\right)\right)\\
=&\frac{2}{K}\sum_{k=1}^{K}\left(1-\prod_{t=1}^{T}\prod_{\substack{l=1\\l\neq k}}^{K}\left(1-\sum_{\alpha=\alpha_0}^{\infty}\Pr\left(\hat{V}_l^{(t)}-\hat{V}_k^{(t)}=\alpha\Big|X=x_k\right)\right)\right)\\
\overset{(e)}{\leq}&\frac{2}{K}\sum_{k=1}^{K}\left(1-\prod_{t=1}^{T}\prod_{\substack{l=1\\l\neq k}}^{K}\left(1-\sum_{\alpha=\alpha_0}^{\infty}e^{-r_tM\left(q_{k|k}^{(t)}+q_{l|k}^{(t)}\right)}\left(\frac{q_{l|k}^{(t)}}{q_{k|k}^{(t)}}\right)^{\frac{\alpha}{2}}I_{|\alpha|}\left(2r_tM\sqrt{q_{l|k}^{(t)}q_{k|k}^{(t)}}\right)\right)\right).
\end{aligned}
\tag{27}
$$

lemma to obtain an upper bound on the error rate of the MVF for i.non-i.d. voters.

Theorem 4: The error rate of the MVF for the case of i.non-i.d. voters is upper bounded by

$$
P_e\leq\frac{2}{K}\sum_{k=1}^{K}\left(1-\prod_{t=1}^{T}\prod_{\substack{l=1\\l\neq k}}^{K}\left(1-\sum_{\alpha=\alpha_0}^{\infty}e^{-r_tM\left(q_{k|k}^{(t)}+q_{l|k}^{(t)}\right)}\left(\frac{q_{l|k}^{(t)}}{q_{k|k}^{(t)}}\right)^{\frac{\alpha}{2}}\times I_{|\alpha|}\left(2r_tM\sqrt{q_{l|k}^{(t)}q_{k|k}^{(t)}}\right)\right)\right),
\tag{26}
$$

where $\alpha_0=r_tM\left(q_{l|k}^{(t)}-q_{k|k}^{(t)}+\delta_{l|k}\right)$.

Proof: According to Lemma 2, (25) needs to hold for at least one $l\in[K]$ and one $t\in[T]$ in order for the MVF to output an incorrect class. Starting from the definition of the error rate, P_e, we use inequality (25) to upper bound P_e as presented in (27), where the right-hand side of (a) in (27) denotes the error rate assuming that the MVF always breaks ties in favor of an incorrect class. Moreover, (b) is a result of Lemma 2 and (c) is a result of Lemma 1. Please note that $\mathbf{V}$ is Multinomial random vector and $\Pr\left(\bigcup_{t=1}^{T}\bigcup_{\substack{l=1\\l\neq k}}^{K}\left\{\hat{V}_l^{(t)}-\hat{V}_k^{(t)}\geq E\left\{V_l^{(t)}-V_k^{(t)}\right\}+r_tM\delta_{l|k}\right\}\Big|X=x_k\right)$ is a monotonic function of M, hence it satisfies the conditions of Lemma 1. Next, (d) is due to the independence of $\hat{V}_l^{(t)}$'s, (d) is a result of Lemma 2, and finally, (e) is derived by substituting $\Pr\left(\hat{V}_l-\hat{V}_k=\alpha|X=x_k\right)$ with the Skellam probability mass function [24]. ■

As we discussed before, a higher $\delta_{l|k}$ is desirable as it shows a lower chance of outputting $\hat{X}=x_l$ while $X=x_k$. This is also evident from (26) since a higher $\delta_{l|k}$ increases the value of α_0 which in turn increases the lower limit of the inner summation. This in turn reduces the right-hand side of (26) as the typical value of the inner summation is always positive.

Next, we simplify the upper bound on the error rate of the MVF for i.non-i.d. voters in order to provide a better intuition about the behavior of (26) versus the number of voters, M, both for the asymptotic and the transient scenarios.

B. Asymptotic Behaviour Of The Error Rate

In the following, we first state a theorem and then exploit it to simplify the upper bound in (26) in the asymptotic scenario.

Theorem 5: Assuming that $\delta_{l_1,l_2|k} > 0$ for given l_1, l_2 and $k \in [K]$, we have

$$\lim_{M\to\infty} \Pr\left(V_{l_1} < V_{l_2} | X = x_k\right) = 0. \tag{28}$$

Proof: Please refer to Appendix E. ∎

Corollary 3: In the asymptotic scenario, where the number of voters goes to infinity, the error rate conditioned on $X = x_k$ satisfies

$$\begin{aligned}\lim_{M\to\infty} P_{e|x_k} &= \lim_{M\to\infty} \Pr\left(\hat{X} = X \middle| X = x_k\right) \\ &= \begin{cases} 0 & \text{if } \forall\ l \neq k, \delta_{l|k} > 0 \\ 1 & \text{if } \exists\ l \neq k, \delta_{l|k} < 0. \end{cases}\end{aligned} \tag{29}$$

Proof: The proof is straightforward. Starting from the first case in (29), if $\delta_{l|k} > 0$, $\forall l \neq k \in [K]$, then due to the Theorem 5, we have

$$\lim_{M\to\infty} \Pr\left(V_k < V_l | X = x_k\right) = 0,\ \forall\, l \neq k \in [K]. \tag{30}$$

which implies that

$$\lim_{M\to\infty} P_{e|x_k} = 0. \tag{31}$$

On the contrary, if we have $\delta_{l|k} < 0$ for at least one $l \neq k \in [K]$, then due to the Theorem 5, we have

$$\lim_{M\to\infty} \Pr\left(V_l < V_k | X = x_k\right) = 0, \tag{32}$$

which implies that

$$\lim_{M\to\infty} P_{e|x_k} = 1. \tag{33}$$

∎

Corollary 4: Let $\mathcal{A}$ be the largest subset of $[K]$ such that $\forall k \in \mathcal{A}$ we have $\delta_{l|k} > 0$, $\forall l \in [K] \setminus k$. Then, the error rate of the MVF-based DFT converges towards $1 - \frac{|\mathcal{A}|}{K}$ as M grows.

Proof: The proof is straightforward. From (15), we have

$$\begin{aligned}\lim_{M\to\infty} P_e &= \lim_{M\to\infty} \sum_{k=1}^{K} \Pr\left(X = x_k\right) P_{e|x_k} \\ &= \sum_{k\in\mathcal{A}} \frac{1}{K} \times 0 + \sum_{k\in[K]\setminus\mathcal{A}} \frac{1}{K} \\ &= 1 - \frac{|\mathcal{A}|}{K}.\end{aligned} \tag{34}$$

∎

Corollary 4 extends Corollary 2, hence generalizing the Condorcet Jury Theorem even further. Due to the Corollary 4, we need $\delta_{l|k} > 0$, $\forall l, k \in [K]$, where $l \neq k$, for the MVF to work reliably, i.e.,

$$\lim_{M\to\infty} P_e = 0 \iff \forall l, k \in [K],\ \text{where } l \neq k,\ \delta_{l|k} > 0. \tag{35}$$

C. Transient Behaviour Of The Error Rate

Assuming $\delta_{l|k} > 0$, $\forall l, k \in [K]$, where $l \neq k$, we can derive an upper bound on the slope of decay of P_e versus M for i.non-i.d. voters, similar to what we had in Theorem 3 for i.i.d. voters. We first define an auxiliary variable, ϵ^*, as

$$\epsilon^* = \min_{\substack{k,l,t \\ k\neq l}} \left(\frac{r_t M \delta_{l|k}^2}{8\left(q_{l|k}^{(t)} + \frac{\delta_{l|k}}{6}\right)} \right). \tag{36}$$

Theorem 6: If $\delta_{l|k} > 0$, $\forall l, k \in [K]$, then we have

$$P_e \leq 2KT \exp\left(-\epsilon^*\right). \tag{37}$$

Proof: Please refer to Appendix F. ∎

The upper bound derived in Theorem 6 is much looser than the upper bound obtained in Theorem 4, however, it is very insightful. The upper bound in (37) shows that the error rate exponentially decays towards 0 as M grows since ϵ^*, defined in (36), increases linearly with M. More specifically, using Theorem 6 we can bound the slope of the decay of the error rate for the case where $\delta_{l|k} > 0$, $\forall l, k \in [K]$, as

$$\frac{\ln P_e}{M} \leq \frac{\ln 2KT}{M} + \frac{\epsilon^*}{M} \approx \min_{\substack{k,l,t \\ k\neq l}} \left(\frac{r_t \delta_{l|k}^2}{8\left(q_{l|k}^{(t)} + \frac{\delta_{l|k}}{6}\right)} \right),\ \text{as } M \to \infty. \tag{38}$$

Therefore, if $\delta_{l|k} > 0$, $\forall l, k \in [K]$, the error rate exponentially decays towards zero with a slope as most $\frac{\ln 2KT}{M} + \frac{\epsilon^*}{M}$, as $M \to \infty$, where for large values of M the upper bound on the slope is a constant number given by $\min\limits_{\substack{k,l,t \\ k\neq l}} \left(\frac{r_t \delta_{l|k}^2}{8\left(q_{l|k}^{(t)} + \frac{\delta_{l|k}}{6}\right)} \right)$.

V. The Majority Voting Function In Comparison With The Truth Discovery Algorithms

The truth discovery algorithms are an important class of DFTs that combine the outputs of several weak classifiers/ voters without ever knowing the ground truth reliability of the classifiers/ voters. While the MVF assumes that all voters are equally reliable, truth discovery algorithms try to guess the reliability index of each voter from their past votes.

In order to obtain a better understanding of the truth discovery algorithms, we will describe it briefly in the following. Let $w_m^{(t)}$ denote the estimated reliability index of voter O_m at the t'th round of running the truth discovery algorithm. The first time that a truth discovery algorithm is executed, when $t = 1$, the truth discovery algorithms often initialize the reliability index of all voters uniformly [15], i.e.,

$$w_m^{(1)} = c, \forall m \in [M]. \tag{39}$$

Once the first decision fusion task is completed, the reliability index of each voter is updated by minimizing a cost function. For instance, consider the cost function $g\left(.,.\right)$ defined in [15]

$$g\left(\mathbf{Y}^{(t)}, \mathbf{w}^{(t)}\right) = \sum_{m=1}^{M} w_m^{(t)} d\left(Y_M^{(t)}, \hat{X^{(t)}}\right), \tag{40}$$

where $\mathbf{w^{(t)}} = \left(w_1^{(t)}, w_2^{(t)}, \ldots, w_M^{(t)}\right)$ is the reliability vector at the beginning of round t, $Y_m^{(t)}$ denotes the value reported

by O_m at round t, $\mathbf{Y^{(t)}} = \left\{Y_1^{(t)}, Y_2^{(t)}, \ldots, Y_M^{(t)}\right\}$, $\hat{X}^{(t)}$ is the estimation of the decision fusion system about the unknown classification, and $d(.,.)$ is a distance function defined by the system design. The goal is to minimize the cost function by optimizing it with respect to $\mathbf{w}^{(t)}$ and $\hat{X}^{(t)}$ recursively.

For a given $k \in [K]$, we compare the reliability of the MVF with the truth discovery algorithms under two alternative assumptions:

1) If $\forall l \neq k \in [K]$ we have $\delta_{l|k} > 0$: We showed in Section III that if $\delta_{l|k} > 0$, $\forall l \neq k \in [K]$, as we increase the number of independent voters, the MVF can obtain an arbitrarily small error rate in detecting X when $X = x_k$. In this scenario, in the best case situation, the truth discovery algorithms may perform even better than the MVF as their performance continues to improve not only by increasing the number of voters but also with time. This is because when $\delta_{l|k} > 0$, $\forall l \neq k \in [K]$, there is a high probability that the final output of the decision unit, $\hat{X}$, at time t, is correct which in turn results in increasing the accuracy of estimation of the reliability index of the voters. Hence, over time the more reliable voters will attract a higher reliability index and help in increasing the accuracy of the decision fusion system. However, this is the best-case situation. There are also circumstances where implementing truth discovery algorithms may result in highly erroneous outcomes even if $\delta_{l|k} > 0$, $\forall l \neq k \in [K]$. There is always a non-zero probability that during the first few rounds of the operation of the decision fusion system, the truth discovery algorithm outputs incorrect values, simply due to the inherent noise of the data sources and processing algorithms that the voters use to prepare their votes. In spite of the lack of ground truth in the system, the truth discovery algorithms often interpret the final output of the system as the ground truth. As a result, the reliability index of the erroneous voters that have voted for the classes that are deemed as correct classes will be increased. Consequently, the error rate of the system increases and may lead to the domination of unreliable voters in the system.

2) If $\exists l \neq k \in [K]$ such that $\delta_{l|k} < 0$: Recalling our results in Section III, if $\exists l \neq k \in [K]$ such that $\delta_{l|k} < 0$, then the MVF will have a high error rate in detecting X when $X = x_k$. In fact, we showed that in this scenario the error rate increases exponentially as we increase the number of independent voters. In this scenario, the performance of the truth discovery algorithms cannot be better that the MVF. From (39), we observe that at time $t = 1$, the output of a typical truth discovery-based DFT is equal to the output of the MVF. If there is at least one $l \in [K]$ such that $\delta_{l|k} < 0$, whether we employ the MVF or the truth discovery algorithms, the output of the data fusion system is erroneous with a high probability[2]. However, in the truth discovery algorithms, those voters that have mistakenly reported the same value as the majority of the voters reported in the first round will receive a higher reliability index at the beginning of the second round. Therefore, those unreliable voters will have a higher influence on the data fusion system in the second round.

In other words, the idea of guessing the reliability of the voters from their past votes without having access to the past ground truths, and then, combining those past votes to generate reliability indexes, creates a positive feedback loop in the system. When compared to the MVF, the positive feedback loop of the truth discovery algorithms in cases when $\delta_{l|k} > 0$, $\forall l \neq k \in [K]$, amplifies the accuracy in the best-case scenario. However, the same positive feedback loop of the truth discovery algorithms in cases when $\delta_{l|k} < 0$, for at least one $l \neq k \in [K]$, amplifies the error rate. In other words, in terms of their error rate, the truth discovery algorithms, in the best case, are simply a MVF with an amplification factor.

Furthermore, the MVF is faster and simpler than the truth discovery algorithms, and since it does not require past votes, it is practical for a wider range of applications. For instance, truth discovery algorithms are not applicable if a decision fusion system requires the voters to join the network anonymously in order to preserve their privacy and neutralize bribing attacks [25]. Also, in some applications the voters may rarely participate in a task (e.g., when reporting an incident in a system) or they may prefer to remain anonymous for security and privacy reasons (e.g., when reporting corruption in an organization), and as a result, zero or little information about their previously reported values is known to the decision unit.

Hence, employing the MVF as a DFT could be a better choice especially when a large number of independent voters are available in the system.

VI. Numerical Results

In this section, we present our numerical results in two subsections assuming i.i.d. and i.non-i.d. voters, respectively. In each subsection, we first explain the simulation parameters and then discuss the results.

A. Numerical Analysis Assuming I.I.D. Voters

In this subsection, we explore the performance of the MVF assuming i.i.d. voters. We simulate a 10-class classification problem, i.e., $|\mathcal{X}| = K = 10$ where the probability of misclassification is uniform over all input and output classes in $\mathcal{X}$. Let $\Pr(Y_m = x_k | X = x_k) = \gamma + \frac{1-\gamma}{K}$ and $\Pr(Y_m = x_l | X = x_k) = \frac{1-\gamma}{K}$. Hence, the probability transition matrix for the voters has the following format

$$\mathrm{p} = \begin{bmatrix} \left(\gamma + \frac{1-\gamma}{K}\right) & \frac{1-\gamma}{K} & \cdots & \frac{1-\gamma}{K} \\ \frac{1-\gamma}{K} & \left(\gamma + \frac{1-\gamma}{K}\right) & \cdots & \frac{1-\gamma}{K} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1-\gamma}{K} & \frac{1-\gamma}{K} & \cdots & \left(\gamma + \frac{1-\gamma}{K}\right) \end{bmatrix}, \quad (41)$$

where $\gamma \in [0, 1]$ is the parameter we use to model the average level of accuracy of the voters. This is a standard way for modeling voters in a decision fusion system that is commonly known as the homogeneous Dawid-Skene model [22]. Using this model, we can study the MVF's error rate versus the accuracy of individual voters. For instance, if $\gamma = 0$, we have the special case when we have zero accuracy and the expected

[2]Some truth discovery implementation may use other statistics like median or just randomly pick a value as the output. Though, in the absence of historical data, non of these alternative methods is more reliable than the MVF in obtaining the first output of the system

number of votes submitted for different classes is uniformly distributed regardless of X, while for $\gamma = 1$, we have full accuracy since all of the voters vote for the correct class.

The performance of our upper bound on the error rate given in Theorem 1, is illustrated in Fig. 2 and is compared to the true error rate obtained by simulations. As we can see in Fig. 2, the error rate of the MVF decreases exponentially as the number of voters increases. Assuming that the individual voters have an accuracy parameter $\gamma = 0.3$, a single voter makes a classification error with a probability as high as 0.63. However, our simulation shows that around 30 independent voters together with a MVF, can achieve an error rate less than 10^{-2}. This number reduces to as low as 15 voters if the accuracy parameter is as high as $\gamma = 0.5$.

In Fig. 2, the upper bound we obtained in Theorem 1 is compared to two other upper bounds previously derived in [21], [22] as well. The simulation results show that our upper bound keeps a constant distance from the true error rate obtained by simulations, while the upper bound obtained in [22] diverges from the true error rate as the number of voters, M, increases and the upper bound derived in [21] does not decrease as the number of independent voters increases.

Next, we explore the performance of the upper bound we derived on the slope of decay of the error rate for the MVF. To this end, we first generate a random probability transition matrix that satisfies the right-hand-side of (35) to model the probability transition matrices of M i.i.d. voters (Fig. 3). For the probability transition matrix in Fig. 3, we have

$$k^* = \underset{k}{\operatorname{argmin}} \left(\min_{l} \left(\sqrt{p_{k|k}} - \sqrt{p_{l|k}} \right) \right) = 10,$$
$$l^* = \underset{l}{\operatorname{argmin}} \left(\sqrt{p_{k^*|k^*}} - \sqrt{p_{l|k^*}} \right) = 7.$$

Therefore, applying our result in Lemma 2 to the probability transition matrix in Fig. 3, we have

$$\lim_{M\to\infty} \frac{\ln P_e}{M} \leq -\left(\sqrt{p_{k^*|k^*}} - \sqrt{p_{l^*|k^*}} \right)^2 + \epsilon = -0.0204.$$

Hence, we expect the error rate to exponentially decay toward 0 as M grows, where the slope of the decay is at least -0.0204. Fig. 4 illustrates our upper bound of the error rate as well as our upper bound on the slope of decay of the error rate.

B. Numerical Analysis Assuming I.Non-I.D. Voters

We use the Weather Sentiment data set provided by Amazon Mechanical Turk [26] for the purpose of our numerical analysis on the error rate of the MVF for i.non-i.d voters. We should emphasize that we do not aim to use this data set to test the performance of the MVF, rather, our purpose is to test the accuracy of our upper bound using real-world probability transition matrices.

The data set contains 6000 classifications of the sentiment of 300 tweets, with gold-standard sentiment labels, provided by 110 voters [26]. The cardinality of the classification labels is $|\mathcal{X}| = K = 5$. We assume that given the true value of each classification task, the votes submitted by different voters are jointly independent. In order to estimate the probability

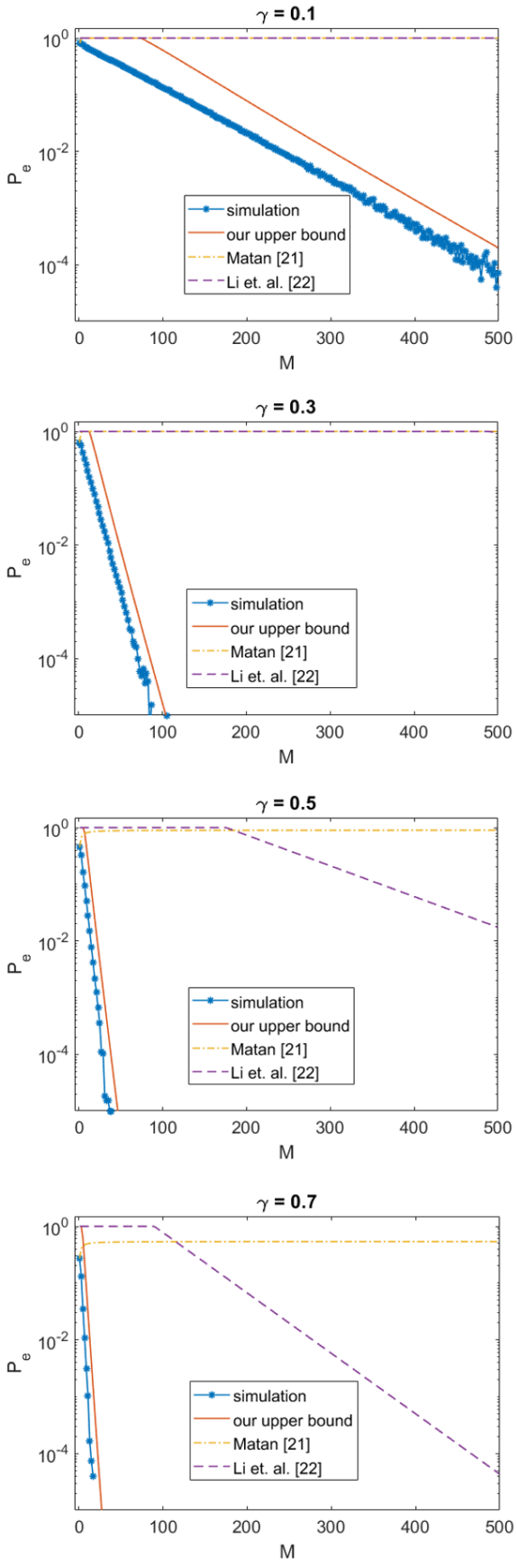

Figure 2. Our theoretical upper bound on the error rate of the majority voting function is plotted for four different values of γ and compared with simulations results and two other upper bounds derived in [22] and [21].

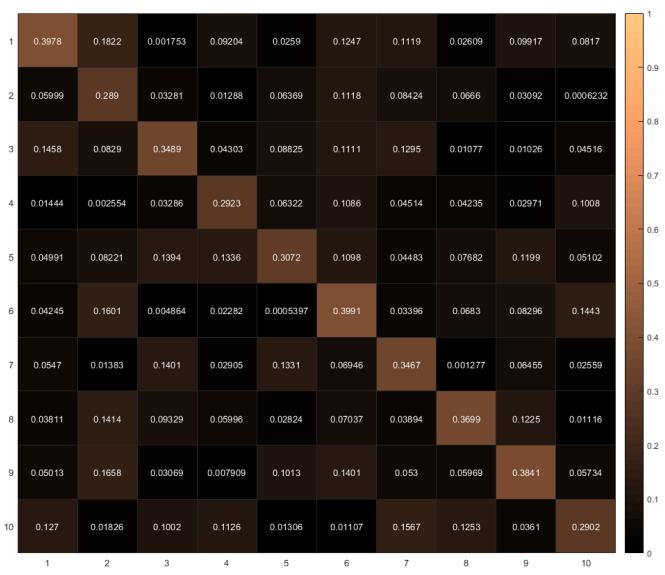

Figure 3. The probability transition matrix of M i.i.d. voters is illustrated. This probability transition matrix is generated randomly and satisfies the right-hand side of (35).

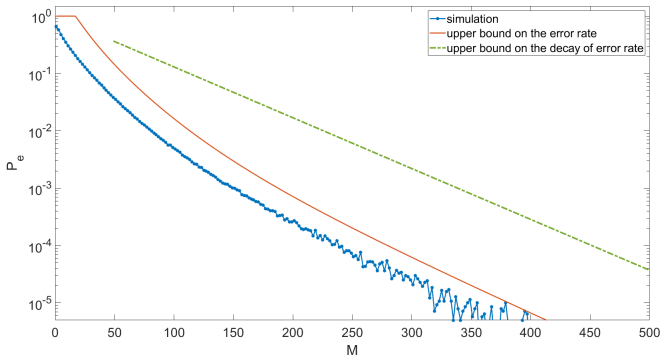

Figure 4. The simulation result, the upper bound, and the upper bound on the slope of decay of the error rate for the MVF are illustrated.

transition matrix of a voter, we need to have enough samples of its reported values in the past. Therefore, we remove the records of all those voters that participated in less than 30 classification tasks. Next, we approximate the probability transition matrix for each one of the 54 remaining voters using the historical data that we have and apply the K-Means algorithm to cluster the voters into six groups based on their probability transition matrices. Finally, we use the probability transition matrices of the centers of each cluster to simulate the probability transition matrices of the voters in 6 partitions, $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_6$. Fig. 5 shows the probability transition matrices used in our simulation. As it is illustrated in Fig. 5, the voters in partition $\mathcal{G}_6$ tend to report class x_3 regardless of the true classification of X. Therefore, one may consider the voters in $\mathcal{G}_6$ as malicious voters.

Using the probability transition matrices in Fig. 5 and assuming that $|\mathcal{G}_1| = 5|\mathcal{G}_2| = 5|\mathcal{G}_3| = 2.5|\mathcal{G}_4| = 2.5|\mathcal{G}_5| = 5|\mathcal{G}_6|$, we obtain negative $\delta_{l|k}$'s $\forall k, l \in [K]$, hence, satisfying the right-hand side of (35). The performance of our upper

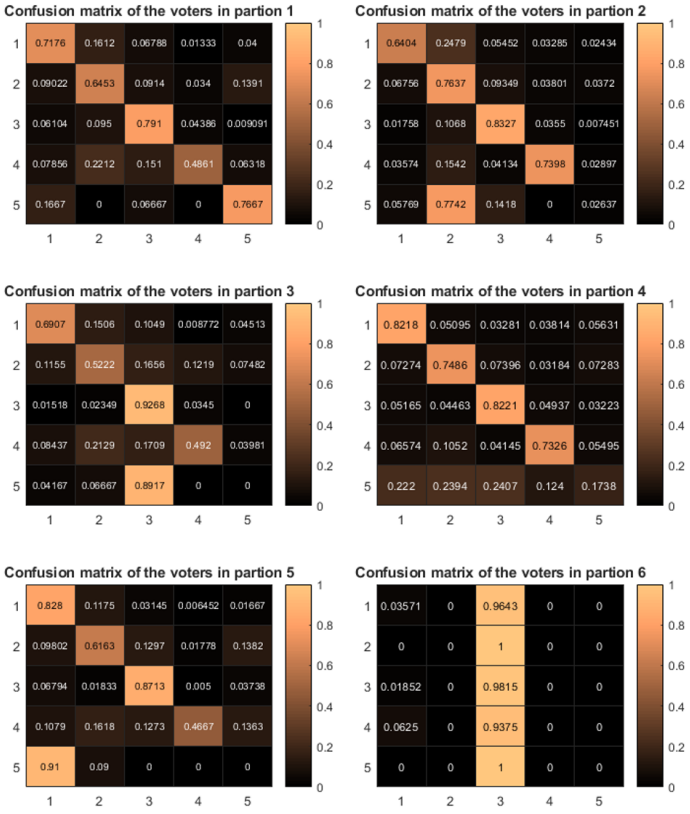

Figure 5. probability transition matrices used for simulating 6 types of voters.

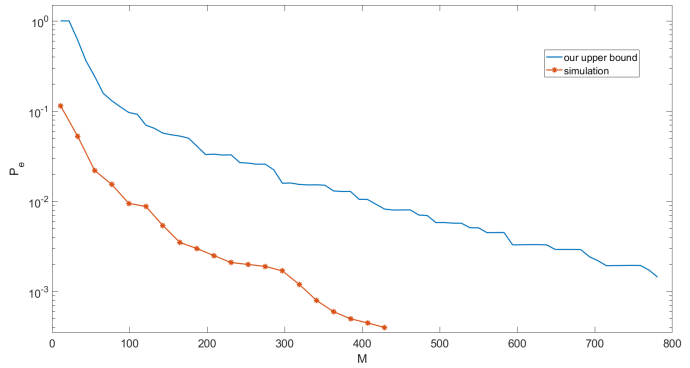

Figure 6. Our upper bound on the error rate of the MVF for i.non-i.d. voters is illustrated and compared to the actual error rate of M independent voters composed of 6 groups with probability transition matrices in Fig. 5 and for group proportions $|\mathcal{G}_1| = 5|\mathcal{G}_2| = 5|\mathcal{G}_3| = 2.5|\mathcal{G}_4| = 2.5|\mathcal{G}_5| = 5|\mathcal{G}_6|$.

bound on the error rate of the MVF for i.non-i.d. voters are illustrated in Fig. 6. As we can see in this figure, both the error rate obtained by simulation as well as the upper bound on error rate derived in Theorem 4 exponentially decay towards 0 as M increases.

VII. Conclusion

In this paper, we have explored the accuracy of the majority voting function. We first assumed a decision fusion system that consists of several independent and identically distributed voters that submit their votes on a classification problem and

then extended our system model to include independent but not necessarily identically distributed voters. We proposed conditions for the accuracy of individual voters and showed that the proposed conditions are necessary and sufficient for achieving a reliable classification using the MVF. More specifically, we proved that when these conditions are met, the error rate goes towards 0 as the number of voters increases towards infinity. On the contrary, if those conditions are not satisfied, then the error rate of the system goes towards 1 as the number of voters grows. We argue that these conditions are also necessary for more complex decision fusion rules such as the weighted majority voting and a range of other fusion rules commonly known as truth discovery algorithms. More importantly, we derived upper bounds on the error rate of the MVF for both i.i.d. and i.non-i.d. voters for general multi-class classification problems. Finally, we presented our numerical results where we compared our upper bounds with the existing upper bounds in the literature and showed the closeness of our upper bounds on the error rate of the MVF to the true error rate obtained by simulation.

REFERENCES

[1] L. I. Kuncheva, *Combining pattern classifiers: methods and algorithms*. John Wiley & Sons, 2014.

[2] L. Breiman, "Bagging predictors," *Machine learning*, vol. 24, no. 2, pp. 123–140, 1996.

[3] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of computer and system sciences*, vol. 55, no. 1, pp. 119–139, 1997.

[4] D. H. Wolpert, "Stacked generalization," *Neural networks*, vol. 5, no. 2, pp. 241–259, 1992.

[5] D. Tao, J. Cheng, Z. Yu, K. Yue, and L. Wang, "Domain-weighted majority voting for crowdsourcing," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 1, pp. 163–174, 2019.

[6] P. Zappi, T. Stiefmeier, E. Farella, D. Roggen, L. Benini, and G. Troster, "Activity recognition from on-body sensors by classifier fusion: sensor scalability and robustness," in *2007 3rd International Conference on Intelligent Sensors, Sensor Networks and Information*, 2007, pp. 281–286.

[7] B. Bigdeli, P. Pahlavani, and H. A. Amirkolaee, "An ensemble deep learning method as data fusion system for remote sensing multisensor classification," *Applied Soft Computing*, vol. 110, p. 107563, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494621004841

[8] J. Adler, R. Berryhill, A. Veneris, Z. Poulos, N. Veira, and A. Kastania, "Astraea: A decentralized blockchain oracle," in *2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, 2018, pp. 1145–1152.

[9] S. Ellis, A. Juels, and S. Nazarov. (2017) Chainlink: A decentralized oracle network. [Online]. Available: https://link.smartcontract.com/whitepaper

[10] L. Lam and C. Suen, "A theoretical analysis of the application of majority voting to pattern recognition," in *Proceedings of the 12th IAPR International Conference on Pattern Recognition, Vol. 3 - Conference C: Signal Processing (Cat. No.94CH3440-5)*, vol. 2, 1994, pp. 418–420 vol.2.

[11] U. G. Mangai, S. Samanta, S. Das, and P. R. Chowdhury, "A survey of decision fusion and feature fusion strategies for pattern classification," *IETE Technical review*, vol. 27, no. 4, pp. 293–307, 2010.

[12] O. Sagi and L. Rokach, "Ensemble learning: A survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 4, p. e1249, 2018.

[13] R. E. Schapire, "The strength of weak learnability," *Machine learning*, vol. 5, no. 2, pp. 197–227, 1990.

[14] S. Eskandari, M. Salehi, W. C. Gu, and J. Clark, *SoK: Oracles from the Ground Truth to Market Manipulation*. New York, NY, USA: Association for Computing Machinery, 2021, p. 127–141. [Online]. Available: https://doi.org/10.1145/3479722.3480994

[15] Q. Li, Y. Li, J. Gao, B. Zhao, W. Fan, and J. Han, "Resolving conflicts in heterogeneous data by truth discovery and source reliability estimation," in *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*, 2014, pp. 1187–1198.

[16] C. Meng, W. Jiang, Y. Li, J. Gao, L. Su, H. Ding, and Y. Cheng, "Truth discovery on crowd sensing of correlated entities," in *Proceedings of the 13th acm conference on embedded networked sensor systems*, 2015, pp. 169–182.

[17] H. Wu, B. Düdder, L. Wang, S. Sun, and G. Xue, "Blockchain-based reliable and privacy-aware crowdsourcing with truth and fairness assurance," *IEEE Internet of Things Journal*, 2021.

[18] K. Nelaturu, J. Adler, M. Merlini, R. Berryhill, N. Veira, Z. Poulos, and A. Veneris, "On public crowdsource-based mechanisms for a decentralized blockchain oracle," *IEEE Transactions on Engineering Management*, vol. 67, no. 4, pp. 1444–1458, 2020.

[19] D. Ruta and B. Gabrys, "A theoretical analysis of the limits of majority voting errors for multiple classifier systems," *Pattern Analysis and Applications*, vol. 5, no. 4, pp. 333–350, 2002.

[20] L. I. Kuncheva, C. J. Whitaker, C. A. Shipp, and R. P. W. Duin, "Limits on the majority vote accuracy in classifier fusion," *Pattern Analysis and Applications*, vol. 6, no. 1, pp. 22–31, 2003.

[21] O. Matan, "On voting ensembles of classifiers (extended abstract)," in *IN PROCEEDINGS OF AAAI-96 WORKSHOP ON INTEGRATING MULTIPLE LEARNED MODELS*, 1996, pp. 84–88.

[22] H. Li and B. Yu, "Error rate bounds and iterative weighted majority voting for crowdsourcing," *arXiv preprint arXiv:1411.4086*, 2014.

[23] M. Mitzenmacher and E. Upfal, *Probability and computing: Randomization and probabilistic techniques in algorithms and data analysis*. Cambridge university press, 2017.

[24] S. JG, "The frequency distribution of the difference between two poisson variates belonging to different populations." *Journal of the Royal Statistical Society. Series A (General)*, vol. 109, no. Pt 3, pp. 296–296, 1946.

[25] Y. Lu, Q. Tang, and G. Wang, "Zebralancer: Private and anonymous crowdsourcing system atop open blockchain," in *2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS)*, 2018, pp. 853–865.

[26] M. VENANZI, W. Teacy, A. Rogers, and N. Jennings, "Weather sentiment - amazon mechanical turk dataset," 2015. [Online]. Available: https://eprints.soton.ac.uk/376543/

[27] N. Cressie, *Statistics for spatial data*. John Wiley & Sons, 2015.

[28] P. Kasperkovitz, "Asymptotic approximations for modified bessel functions," *Journal of Mathematical Physics*, vol. 21, no. 1, pp. 6–13, 1980.

[29] T. J. I. Bromwich *et al.*, *An introduction to the theory of infinite series*. American Mathematical Soc., 2005, vol. 335.

[30] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. USA: Wiley-Interscience, 2006.

[31] F. Chung and L. Lu, "Concentration inequalities and martingale inequalities: a survey," *Internet mathematics*, vol. 3, no. 1, pp. 79–127, 2006.

APPENDIX A
PROOF OF LEMMA 1

In order to prove Lemma 1, we use a bound proposed in Theorem 5.10 in [23]. This bound is provided below as a lemma for completeness.

Lemma 3: Let $\mathbf{V} = (V_1, V_2, \ldots, V_K)$ be a random vector with the following Multinomial distribution

$$\Pr(\mathbf{V} = \mathbf{v} | X = x_k) = \frac{M!}{v_1! v_2! \ldots v_K!} \times p_{1|k}^{v_1} \times p_{2|k}^{v_2} \times \cdots \times p_{K|k}^{v_K}, \tag{42}$$

such that $\sum_{k=1}^{K} v_k = M$ and

$$p_{1|k} = p_{2|k} = \cdots = p_{K|k} = \frac{1}{K}, \forall k \in [K]. \tag{43}$$

Moreover, let $\hat{\mathbf{V}} = \left(\hat{V}_1, \hat{V}_2, \ldots, \hat{V}_K\right)$ be a vector of independent and identically distributed Poisson RVs with parameter $\lambda = \frac{M}{K}$. Finally, let $\mathcal{E}(\mathbf{V})$ be an event imposed on the Multinomial random vector $\mathbf{V}$ whose probability is either

monotonically increasing or monotonically decreasing in the number of voters, M, and let $\mathcal{E}\left(\hat{\mathbf{V}}\right)$ denote the same event applied on the random Poisson vector $\hat{\mathbf{V}}$. Then, we have

$$\Pr\left(\mathcal{E}\left(\mathbf{V}\right)\right) \leq 2\Pr\left(\mathcal{E}\left(\hat{\mathbf{V}}\right)\right). \tag{44}$$

Proof: The proof[3] is provided in [23]. ∎

In Lemma 3, it is assumed that (43) holds. In order to prove Lemma 1, we need to expand Lemma 3 to the more general distribution of $p_{l|k}$'s where the conditional probabilities are not necessarily as in (43), i.e., uniformly distributed. To this end, we first define the set $\mathcal{Z} = \{z_1, z_2, \ldots, z_R\}$ where $R \to \infty$. Next, for each x_k we define the many-to-one mapping, $f_k : \mathcal{Z} \to \mathcal{X}$, as

$$f_k\left(z_r\right) = \begin{cases} x_1 & r \leq Rp_{1|k}, \\ x_l & R\sum_{i=1}^{l-1} p_{i|k} < r \leq R\sum_{i=1}^{l} p_{i|k}. \end{cases} \tag{45}$$

In words, using f_k, we group the first $Rp_{1|k}$ elements in $\mathcal{Z}$ and map it to class x_1. Similarly, we group the next $Rp_{2|k}$ elements in $\mathcal{Z}$ and map it to class x_2, and so on until the last $Rp_{K|k}$ elements in $\mathcal{Z}$ that are mapped to the last class in $\mathcal{X}$, i.e., x_K.

Given that $X = x_k$, let $g_k(.)$ be the random oracle that reads any submitted vote, Y_m, and maps it to an element in its pre-image, $f_k^{-1}\left(Y_m\right)$, uniformly at random. Moreover, let $\mathbf{W} = (W_1, W_2, \ldots, W_R)$, where W_r denotes the number of votes mapped to z_r, we have

$$\Pr\left(\mathbf{W} = \mathbf{w} | X = x_k\right) = \frac{M!}{w_1! w_2! \ldots w_R!} \prod_{r=1}^{R} \left(\frac{1}{R}\right)^{w_r}, \tag{46}$$

where $\sum_{r=1}^{R} w_r = M$. This is similar to what we had in Lemma 3 and is resulted from the fact the probability of any randomly selected vote gets mapped to z_r is equal for all $r \in [R]$, i.e.,

$$\Pr\left(g_k\left(f_k^{-1}\left(Y_m\right)\right) = z_r\right) = \frac{1}{R}. \tag{47}$$

Furthermore, let $\hat{\mathbf{W}} = \left(\hat{W}_1, \hat{W}_2, \ldots, \hat{W}_R\right)$, where $\hat{W}_r$ denotes the number of votes mapped to z_r in the case where the total number of voters is assumed to be a Poisson random variable with parameter M. By exploiting the thinning property of Poisson RVs [27], we have

$$\begin{aligned} \Pr\left(\hat{\mathbf{W}} = \hat{\mathbf{w}} | X = x_k\right) &= \prod_{l=1}^{K} \Pr\left(\hat{W}_r = \hat{w}_r | X = x_k\right) \\ &= \prod_{r=1}^{R} \frac{e^{-\frac{M}{R}} \left(\frac{M}{R}\right)^{\hat{w}_r}}{\hat{w}_r!}, \end{aligned} \tag{48}$$

which shows that $\hat{W}_r$'s are independent Poisson RVs. We can use Lemma 3 to show that for any event $\mathcal{E}$ whose probability is monotonically increasing or decreasing with M, we have

$$\Pr\left(\mathcal{E}\left(\mathbf{W}\right)\right) \leq 2\Pr\left(\mathcal{E}\left(\hat{\mathbf{W}}\right)\right). \tag{49}$$

[3]The original Theorem has been stated for the balls and bins problem. However, we presented Lemma 3 for voters and classes in a classification problem in order to keep our terminology consistent.

Recalling that the number of votes submitted for x_l, V_l, is given by

$$V_l = \sum_{r \in f_k^{-1}(x_1)} W_r. \tag{50}$$

Therefore, any event on $\mathbf{V}$ (or $\hat{\mathbf{V}}$) is uniquely mapped to an event on $\mathbf{V}$ (or $\hat{\mathbf{V}}$). More specifically,

$$\begin{aligned} \Pr\left(\mathcal{E}\left(V_1, V_2, \ldots, V_K\right)\right) &= \Pr\left(\mathcal{E}\left(\textstyle\sum_{r \in f_k^{-1}(x_1)} W_r, \ldots, \sum_{r \in f_k^{-1}(x_K)} W_r\right)\right) \\ &\leq 2\Pr\left(\mathcal{E}\left(\textstyle\sum_{r \in f_k^{-1}(x_1)} \hat{W}_r, \ldots, \sum_{r \in f_k^{-1}(x_K)} \hat{W}_r\right)\right) \\ &\overset{(a)}{=} 2\Pr\left(\mathcal{E}\left(\hat{V}_1, \hat{V}_2, \ldots, \hat{V}_K\right)\right), \end{aligned} \tag{51}$$

where (a) is obtained from the fact that the sum of independent Poisson RVs is Poisson. Therefore, the proof is complete.

Appendix B
Proof Of Theorem 2

Using Lemma 1, we bound $\Pr\left(V_{l_1} \leq V_{l_2} | X = x_k\right)$ as

$$\Pr\left(V_{l_1} \leq V_{l_2} | X = x_k\right) \leq 2\Pr\left(\hat{V}_{l_1} \leq \hat{V}_{l_2} | X = x_k\right), \tag{52}$$

where $\hat{V}_{l_1}$ and $\hat{V}_{l_2}$ are independent Poisson RVs with parameter $\lambda_{l_1|k} = Mp_{l_1|k}$ and $\lambda_{l_2|k} = Mp_{l_2|k}$, respectively. Next, we use Chernoff bound to derive an upper bound on $\Pr\left(\hat{V}_{l_1} \leq \hat{V}_{l_2} | X = x_k\right)$, as

$$\begin{aligned} \Pr\left(\hat{V}_{l_1} \leq \hat{V}_{l_2} | X = x_k\right) &\leq E\left\{e^{t\left(\hat{V}_{l_2} - \hat{V}_{l_1}\right)}\right\} \\ &= E\left\{e^{t\hat{V}_{l_2}}\right\} E\left\{e^{-t\hat{V}_{l_1}}\right\} \\ &= e^{\lambda_{l_2|k}\left(e^t - 1\right) + \lambda_{l_1|k}\left(e^{-t} - 1\right)}. \end{aligned} \tag{53}$$

By taking the derivative of right hand side of (53) and solving it for t, we find t_0, which is the value of t for $t > 0$ that minimizes (52), as

$$t_0 = \ln \frac{\sqrt{\lambda_{l_2|k}\lambda_{l_1|k}}}{\lambda_{l_2|k}}. \tag{54}$$

Please note that $p_{l_1|k} > p_{l_2|k}$, hence, $t_0 = \ln \frac{\sqrt{\lambda_{l_2|k}\lambda_{l_1|k}}}{\lambda_{l_2|k}}$ is a positive and satisfies $t > 0$.

By substituting t in (53) with t_0 in (54), we have for (53) the following

$$\begin{aligned} \Pr\left(V_{l_1} \leq V_{l_2} | X = x_k\right) &\leq 2\Pr\left(\hat{V}_{l_1} \leq \hat{V}_{l_2} | X = x_k\right) \\ &\leq 2e^{-\left(\sqrt{\lambda_{l_1|k}} - \sqrt{\lambda_{l_2|k}}\right)^2} \\ &= 2e^{-M\left(\sqrt{p_{l_1|k}} - \sqrt{p_{l_2|k}}\right)^2}. \end{aligned} \tag{55}$$

Appendix C
Proof Of Theorem 3

We can bound the modified Bessel function of the first kind, $I_{|\alpha|}\left(x\right)$, as [28]

$$I_{|\alpha|}\left(x\right) = \frac{e^x e^{\frac{-\alpha^2}{x}}}{\sqrt{2\pi x}}\left(1 + \epsilon_0\right) \leq \frac{e^x}{\sqrt{2\pi x}}\left(1 + \epsilon_0\right), \tag{56}$$

$$P_e \leq 2\left(1-\frac{1}{K}\sum_{k=1}^{K}\prod_{\substack{l=1\\ l\neq k}}^{K}\left(1-\sum_{\alpha=0}^{\infty}e^{-M\left(p_{k|k}+p_{l|k}\right)}\left(\frac{p_{l|k}}{p_{k|k}}\right)^{\frac{\alpha}{2}}\frac{e^{\left(2M\sqrt{p_{l|k}p_{k|k}}\right)}\left(1+\epsilon_0\right)}{\sqrt{2\pi 2M\sqrt{p_{l|k}p_{k|k}}}}\right)\right)$$
$$=2\left(1-\frac{1}{K}\sum_{k=1}^{K}\prod_{\substack{l=1\\ l\neq k}}^{K}\left(1-\frac{e^{-M\left(\sqrt{p_{k|k}}-\sqrt{p_{l|k}}\right)^2}\left(1+\epsilon_0\right)}{\sqrt{2\pi 2M\sqrt{p_{l|k}p_{k|k}}}}\sum_{\alpha=0}^{\infty}\left(\frac{p_{l|k}}{p_{k|k}}\right)^{\frac{\alpha}{2}}\right)\right)$$
$$\overset{(a)}{=}2\left(1-\frac{1}{K}\sum_{k=1}^{K}\prod_{\substack{l=1\\ l\neq k}}^{K}\left(1-\frac{e^{-M\left(\sqrt{p_{k|k}}-\sqrt{p_{l|k}}\right)^2}\left(1+\epsilon_0\right)}{\sqrt{2\pi 2M\sqrt{p_{l|k}p_{k|k}}}\left(1-\sqrt{\frac{p_{l|k}}{p_{k|k}}}\right)}\right)\right). \tag{58}$$

where $|\epsilon_0|$ is of order $x^{\frac{-1}{4}}$ [28]. Therefore, we have

$$I_{|\alpha|}\left(2M\sqrt{p_{l|k}p_{k|k}}\right)\leq\frac{e^{2M\sqrt{p_{l|k}p_{k|k}}}\left(1+\epsilon_0\right)}{\sqrt{2\pi 2M\sqrt{p_{l|k}p_{k|k}}}}. \tag{57}$$

By using (56) to substitute $I_{|\alpha|}\left(2M\sqrt{p_{l|k}p_{k|k}}\right)$ in (9), we obtain the inequality in (58), where (a) in (58) follows from the following equality

$$\sum_{\alpha=0}^{\infty}\left(\frac{p_{l|k}}{p_{k|k}}\right)^{\frac{\alpha}{2}}=\frac{1}{1-\sqrt{\frac{p_{l|k}}{p_{k|k}}}}.$$

We upper bound the right-hand-side of (58) to obtain

$$P_e\overset{(b)}{\leq}2\left(1-\prod_{\substack{l=1\\ l\neq k^*}}^{K}\left(1-\frac{e^{-M\left(\sqrt{p_{k^*|k^*}}-\sqrt{p_{l|k^*}}\right)^2}\left(1+\epsilon_0\right)}{\sqrt{2\pi 2M\sqrt{p_{l|k^*}p_{k^*|k^*}}}\left(1-\sqrt{\frac{p_{l|k^*}}{p_{k^*|k^*}}}\right)}\right)\right)$$
$$\overset{(c)}{\leq}2\left(\sum_{\substack{l=1\\ l\neq k^*}}^{K}\frac{e^{-M\left(\sqrt{p_{k^*|k^*}}-\sqrt{p_{l|k^*}}\right)^2}\left(1+\epsilon_0\right)}{\sqrt{4\pi M\sqrt{p_{l|k^*}p_{k^*|k^*}}}\left(1-\sqrt{\frac{p_{l|k^*}}{p_{k^*|k^*}}}\right)}\right), \tag{59}$$

where (b) follows from the fact that substituting k with k^* in (58) maximizes the right-hand-side of the inequality and (c) is obtained using the following inequality. Let $a_1, a_2, \ldots, a_N$ be real numbers in $[0,1]$, then we have

$$1-\prod_{n=1}^{N}\left(1-a_n\right)\leq\sum_{n=1}^{N}a_n.$$

This inequality is known as Weierstrass Inequality and can be proved using induction [29].

Next, we upper bound the right-hand-side of (59) to obtain

$$P_e\leq 2\left(K-1\right)\frac{e^{-M\left(\sqrt{p_{k^*|k^*}}-\sqrt{p_{l^*|k^*}}\right)^2}\left(1+\epsilon_0\right)}{\sqrt{4\pi M\sqrt{p_{l^*|k^*}p_{k^*|k^*}}}\left(1-\sqrt{\frac{p_{l^*|k^*}}{p_{k^*|k^*}}}\right)}, \tag{60}$$

where this upper bound follows from the fact that substituting l by l^* in the typical element of the summation maximizes the right-hand-side of (59). Finally, we take the logarithm of both sides and divide the result by M to obtain the slope of the decay of error rate versus the number of voters as

$$\lim_{M\to\infty}\tfrac{\ln P_e}{M}\leq\lim_{M\to\infty}\tfrac{1}{M}\left(\ln\gamma-M\left(\sqrt{p_{k^*|k^*}}-\sqrt{p_{l^*|k^*}}\right)^2-\tfrac{1}{2}\ln M\right), \tag{61}$$

where $\gamma=\frac{2(K-1)(1+\epsilon_0)}{\sqrt{4\pi\sqrt{p_{l|k^*}p_{k^*|k^*}}}\left(1-\sqrt{\frac{p_{l|k^*}}{p_{k^*|k^*}}}\right)}$. Therefore,

$$\lim_{M\to\infty}\frac{\ln P_e}{M}\leq-\left(\sqrt{p_{k^*|k^*}}-\sqrt{p_{l^*|k^*}}\right)^2+\epsilon, \tag{62}$$

where $\epsilon=\lim_{M\to\infty}\frac{1}{M}\left(\ln\gamma-\frac{1}{2}\ln M\right)=0$, since $O\left(\epsilon\right)=\frac{O(\epsilon_0)-\ln\left(M^{-1/2}\right)}{M}=\frac{\ln\left(M^{-3/4}\right)}{M}$. Hence, the proof is complete.

Appendix D
Proof of Lemma 2

We prove Lemma 2 through contradiction. To this end, without loss of generality, let's assume that the MVF outputs class x_{l^*} where $x_{l^*}\in\mathcal{X}$, $x_{l^*}\neq x_k$, and concurrently we have

$$v_l^{(t)}-v_k^{(t)}<E\{V_l^{(t)}-V_k^{(t)}|X=x_k\}+r_tM\delta_{l|k},\quad\forall\, l\in[K]\text{ and }\forall\, t\in[T]. \tag{63}$$

Since (63) holds $\forall l\in[K]$ and $\forall t\in[T]$, we can take the summation of both sides of (63) with respect to t and obtain

$$\sum_{t=1}^{T}\left(v_l^{(t)}-v_k^{(t)}\right)<\sum_{t=1}^{T}\left(E\{V_l^{(t)}-V_k^{(t)}|X=x_k\}+r_tM\delta_{l|k}\right)$$
$$=\sum_{t=1}^{T}E\{V_l^{(t)}-V_k^{(t)}|X=x_k\}+\textstyle\sum_{t=1}^{T}r_tM\delta_{l|k}$$
$$\overset{(a)}{=}\sum_{t=1}^{T}\left(r_tMq_{l|k}^{(t)}-r_tMq_{k|k}^{(t)}\right)+M\delta_{l|k}\textstyle\sum_{t=1}^{T}r_t$$
$$=M\sum_{t=1}^{T}r_t\left(q_{l|k}^{(t)}-q_{k|k}^{(t)}\right)+M\delta_{l|k}$$
$$\overset{(b)}{=}0, \tag{64}$$

where (a) is obtained using the fact that $\mathbf{V}^{(t)}$ is a Multinomial random vector, and as a result, $E\{V_l^{(t)}|X=x_k\}=r_tMq_{l|k}^{(t)}$, $\forall l,k\in[K]$. Next, (b) is obtained by substituting $\delta_{l|k}$ from (24). Therefore, (64) shows that $\forall l\in[K]$, including l^*, we have

$$v_l-v_k=\sum_{t=1}^{T}\left(V_l^{(t)}-V_k^{(t)}\right)<0,\ \forall l\in[K], \tag{65}$$

which contradicts the assumption that the MVF incorrectly outputs class x_{l^*}. Therefore, the inequality in (63) cannot be satisfied and the proof is complete.

$$\begin{aligned}
\lim_{M\to\infty} \Pr\left(V_{l_1} < V_{l_2} | X = x_k\right) &\overset{(a)}{\leq} \lim_{M\to\infty} \Pr\left(\bigcup_{t=1}^{T} V_{l_2}^{(t)} - V_{l_1}^{(t)} \geq E\{V_{l_2}^{(t)} - V_{l_1}^{(t)}\} + Mr_t\delta_{l_1,l_2|k} | X = x_k\right) \\
&\overset{(b)}{\leq} \lim_{M\to\infty} \sum_{t=1}^{T} \Pr\left(V_{l_2}^{(t)} - V_{l_1}^{(t)} \geq E\{V_{l_2}^{(t)} - V_{l_1}^{(t)}\} + Mr_t\delta_{l_1,l_2|k} | X = x_k\right) \\
&\overset{(c)}{\leq} \lim_{M\to\infty} \sum_{t=1}^{T} \Bigg(\Pr\left(V_{l_2}^{(t)} - V_{l_1}^{(t)} - E\{V_{l_2}^{(t)} - V_{l_1}^{(t)}\} \geq Mr_t\delta_{l_1,l_2|k} | X = x_k\right) \\
&\qquad + \Pr\left(V_{l_2}^{(t)} - V_{l_1}^{(t)} - E\{V_{l_2}^{(t)} - V_{l_1}^{(t)}\} \leq -Mr_t\delta_{l_1,l_2|k} | X = x_k\right) \Bigg) \\
&\overset{(d)}{=} \sum_{t=1}^{T} \lim_{M\to\infty} \Pr\left(|V_{l_2}^{(t)} - V_{l_1}^{(t)} - E\{V_{l_2}^{(t)} - V_{l_1}^{(t)}\}| \geq Mr_t\delta_{l_1,l_2|k} | X = x_k\right) \\
&\overset{(e)}{=} 0,
\end{aligned} \tag{66}$$

$$\begin{aligned}
Pr\left(\hat{X} \neq X\right) &\overset{(a)}{=} \frac{1}{K}\sum_{k=1}^{K}\left(1 - \prod_{l=1}^{K}\left(1 - \Pr\left(\hat{X} = x_l | X = x_k\right)\right)\right) \\
&\overset{(b)}{=} \frac{1}{K}\sum_{k=1}^{K}\left(1 - \prod_{l=1}^{K}\prod_{t=1}^{T}\left(1 - \Pr\left(V_l^{(t)} \geq E\{V_l^{(t)}\} + \frac{r_tM\delta_{l|k}}{2}\right)\right)\left(1 - \Pr\left(V_k^{(t)} \leq E\{V_k^{(t)}\} - \frac{r_tM\delta_{l|k}}{2}\right)\right)\right) \\
&\overset{(c)}{\leq} \frac{1}{K}\sum_{k=1}^{K}\left(1 - \prod_{l=1}^{K}\prod_{t=1}^{T}\left(1 - \exp\left(\frac{-\left(r_tM\right)^2\delta_{l|k}^2}{8\left(E\{V_l^{(t)}\} + \frac{r_tM\delta_{l|k}}{6}\right)}\right)\right)\left(1 - \exp\left(\frac{-\left(r_tM\right)^2\delta_{l|k}^2}{8E\{V_k^{(t)}\}}\right)\right)\right) \\
&\overset{(d)}{=} \frac{1}{K}\sum_{k=1}^{K}\left(1 - \prod_{l=1}^{K}\prod_{t=1}^{T}\left(1 - \exp\left(\frac{-r_tM\delta_{l|k}^2}{8\left(q_{l|K}^{(t)} + \frac{\delta_{l|k}}{6}\right)}\right)\right)\left(1 - \exp\left(\frac{-r_tM\delta_{l|k}^2}{8q_{k|k}^{(t)}}\right)\right)\right) \\
&\overset{(e)}{\leq} \frac{1}{K}\sum_{k=1}^{K}\left(\sum_{l=1}^{K}\sum_{t=1}^{T}\left(\exp\left(\frac{-r_tM\delta_{l|k}^2}{8\left(q_{l|K}^{(t)} + \frac{\delta_{l|k}}{6}\right)}\right) + \exp\left(\frac{-r_tM\delta_{l|k}^2}{8q_{k|k}^{(t)}}\right)\right)\right) \\
&\overset{(f)}{\leq} 2KT\left(\exp\left(-\epsilon^*\right)\right),
\end{aligned} \tag{67}$$

Appendix E
Proof Of Theorem 5

To prove Theorem 5, we bound $\lim_{M\to\infty} \Pr\left(V_{l_1} < V_{l_2} | X = x_k\right)$ as presented in (66), where (a) in (66) is a result of Lemma 2, (b) is obtained using the union bound technique, (c) is derived by adding a positive value to the typical element of the summation in (b), and (d) is resulted from combining the two probability expressions in (c) into one probability expression. Finally, (e) is due to the weak law of large numbers [30] used to bound the difference of RV $V_{l_1} - V_{l_2}$ and its mean value.

Appendix F
Proof Of Theorem 6

We bound the error rate as presented in (67), where (a) in (67) is obtained from the definition of error rate. Moreover, (b) follows from Lemma 2 which states that if $\hat{X} \neq X$ while $X = x_k$, there exists at least one $l \neq k \in [K]$ such that

$$V_l^{(t)} - V_k^{(t)} \geq E\{V_l^{(t)} - V_k^{(t)}\} + r_tM\delta_{l|k}. \tag{68}$$

If inequality (68) holds, at least one of the following inequalities must hold

$$\begin{cases} V_l^{(t)} \geq E\{V_l^{(t)}\} + \frac{r_tM\delta_{l|k}}{2}, \\ V_k^{(t)} \leq E\{V_k^{(t)}\} - \frac{r_tM\delta_{l|k}}{2}. \end{cases} \tag{69}$$

Therefore, (b) is a result of (69). Additionally, (c) is obtained using Chernoff bound on the upper and lower tail of Binomial RVs [31], $V_{l|k}^{(t)}$ and $V_{k|k}^{(t)}$, (d) is resulted from substituting $E\{V_l^{(t)}\}$ and $E\{V_k^{(t)}\}$ by $r_tMq_{l|k}^{(t)}$ and $r_tMq_{k|k}^{(t)}$, respectively. Furthermore, (e) is a result of Weierstrass Inequality induction [29] presented in (60), and (f) is obtained using

$$\exp\left(\frac{-r_tM\delta_{l|k}^2}{8\left(q_{l|K}^{(t)} + \frac{\delta_{l|k}}{6}\right)}\right) + \exp\left(\frac{-r_tM\delta_{l|k}^2}{8q_{k|k}^{(t)}}\right) \leq 2\exp\left(-\epsilon^*\right), \quad \forall\, l,k \in [K], \text{ and } \forall\, t \in [T]. \tag{70}$$