Machine Learning Misclassification Within Status Update Optimization

Alessandro Buratto*, Begüm Yivli†, and Leonardo Badia*
* Dept. of Information Engineering (DEI), University of Padova, Italy
email: {alessandro.buratto.1@phd. , leonardo.badia@}unipd.it
† Dept. of Computer Engineering, Boğaziçi University, Türkiye
email: begum.yivli@boun.edu.tr

***Abstract*—This paper explores the optimization of status updates in sensing systems, focusing on misclassification in machine learning (ML) models. Previous research has primarily tackled the impact of different techniques throughout the communication layers on Age of Information (AoI), or alternatively studied the Age of Incorrect Information (AoII) as a flaw that can be counteracted by a more active transmission pattern. Our study presents analytical considerations, as well as simulation results from real datasets, with the original aspect that classification errors are not an externality, but are triggered by a fraction of the status updates themselves, which therefore ought to be kept under control. An excessively high number of transmission may be damaging the system, and the right balance needs to be found between prompt updating that lowers AoI, and accuracy to minimize AoII at the same time. In this sense, we offer a new standpoint for timely status update, where freshness of correct information is required for smart systems to make the best decision in real-time.**

***Index Terms*—Age of information; Age of incorrect information; Machine learning; Cross-layer optimization; Data Acquisition.**

I. Introduction

Nowadays, the need for up-to-date information is undeniable in the ever-changing landscape of smart devices and interconnected systems [1], [2]. However, the importance of data accuracy is equally essential. This duality highlights two key aspects that must be taken into account in modern data-driven applications: the freshness of the data and its correctness [3].

To illustrate this, consider the example of tracking a vehicle's location using a combination of GPS and WiFi technologies [4], [5]. Both obsolete position reporting and wrong estimates can lead to navigation errors or misguided decision-making with fatal consequences [6]. Similarly, in the healthcare sector, monitoring a patient's vital signs and health parameters is heavily reliant on data timeliness and accuracy [7]. Stale data can cause overlook a serious health issue that requires immediate attention, but at the same time a rushed wrong classification may trigger unnecessary alarms and interventions, causing undue stress and potentially compromising patient care [8], [9].

Historically, the main focus in data-driven systems, particularly in the Internet of Things (IoT) domain, has been on the timeliness of data updates. Much research has been done to address the challenge of minimizing the Age of Information (AoI) through efficient scheduling of status updates [10]–[13]. This metric measures the time elapsed since the last successful update of a particular piece of information. By minimizing AoI, IoT systems strive to ensure that decision-making processes are based on the most current available data.

Although the pursuit of minimizing AoI is still highly relevant and critical, there is an increasing recognition that the correctness of data is just as important. Timely but inaccurate data can be as damaging as outdated information [14]. Changing focus from emphasizing data freshness to incorporating data correctness marks a significant evolution for data-driven systems [15].

The scope of our research endeavor is on timely status updates within the context of a Machine Learning (ML) model classifying incoming data. We recognize that, due to the inherent imperfections of ML models, there exists a non-zero probability of misclassification [16], which can lead to error amplification and/or making wrong control decisions. Consequently, it is imperative to address not only the timeliness of data updates, but also the quality and correctness of the information [17].

Thus, our primary focus shifts toward including also a similar metric known as the Age of Incorrect Information (AoII) [15], [18]. Unlike the conventional AoI metric, which emphasizes the freshness of data, AoII places significant emphasis on the potential consequences of delivering incorrect or inaccurate data [3].

In essence, AoII penalizes data updates that result in incorrect classifications by the ML model. It encapsulates the detrimental effects of misinformation and provides a holistic perspective on the cost associated with erroneous data. By shifting our attention to AoII, we recognize that the consequences of misclassification can be severe, ranging from misinformed decisions in critical applications to compromised system reliability in various domains [18].

Our analytical formulation considers a sensor sending status updates to a receiver over a finite time horizon, driven by a penalty function to be minimized, being the linear combination of AoI and AoII through a variable weight [19]. The problem resembles that of scheduling a constrained number of transmission updates over an erasure channel [11], [15], [20], with the important difference that in that scenario, the only impairment is due to the channel which causes the information to be less accurate, but this can be counteracted by retransmissions [21] or additional transmissions [22].

Conversely, we argue that achieving minimal penalty in the case of imperfect classification (as opposed to just erasures) lies in a balanced choice of the transmissions to perform in the considered time horizon, since too many status updates can harm the problem's objective when they do not bring significantly fresh information and at the same time imply a higher risk of misclassification that increases AoII instead. This trend becomes even more acute if the weight of AoII in the linear combination increases, which is also a key parameter that we are going to investigate [23].

We further apply our framework to a real case scenario of a support vector machine (SVM), whose 1-accuracy score is used to derive the probability of misclassification [24]. Finally, we show how the most critical scenario is when few updates can be sent and the impact of wrong classification is more severe, since in this case it is difficult to correct classification errors. Such a trend eventually flattens out as the number of updates increases.

The remainder of the paper is subdivided as follows. In Section II, we discuss the state of the art relevant to our work. In Section III, we describe the analysis of our scenario. In Section IV, we present the ML model we used for our experiments and its use pipeline. In Section V we discuss the results obtained from our numerical evaluations and finally in Section VI we make the final remarks.

II. Related Work

In recent research, numerous academic papers are delving into the concept of AoI within communication systems, with a particular emphasis on its relevance to remote sensing applications in the context of the Internet of Things (IoT) [10], [25], [26]. One noteworthy group of studies focuses on the theoretical assessment of AoI concerning various medium access policies and queueing strategies [27]. This constitutes a fundamental area of investigation, but somewhat distinct from our analysis, which concentrates just on a single node transmitting updates [19], [20].

Another frequently encountered scenario involves integrating AoI into the objective function. This typically aims to minimize the average AoI over an extended period at a receiving node while imposing constraints on the average transmission count at the source node [28].

In particular, papers like [11], [20], [22], [29] share the similarity of planning the timing of status updates, referred to as scheduling. In [29], the problem considers multiple sources with independent arrivals sharing a channel. Conversely, [11] focuses on a single source but a broadcast communication. References [20] and [22] study instead a single unicast source transmitting over an error-prone channel, and present different policies for AoI optimization. In the present paper, we take inspiration from the stateless policy presented in the former of these two references.

However, the important difference with all these previous contributions is that we consider not just AoI, but also AoII in the evaluation. This translates the communication issue into semantic aspects, since the content accuracy becomes relevant beyond its age [23]. We actually look at a linear combination of AoI and AoII; more precisely, as will be clear in the following, we consider a baseline penalty related to AoI that is constantly receiving unit weight in all the evaluations, plus an increase of AoII times a multiplicative coefficient denoted as $Z \geq 1$ that is due to wrong classifications following the reception of a status update.

In other words, if the semantics of the communication is ignored and the failure events simply correspond to packet losses, we can set $Z = 1$ to imply that we treat information ageing phenomena independently of whether said information is correct or not. Conversely, for mission-critical assignments where the accuracy of information is relevant, we will evaluate the impact of a higher coefficient Z.

Finally, our aim is also to connect this penalty combining AoI and AoII with the outcome of a machine learning process [1]. Nowadays, smart cyber-physical systems are expected to make real time decisions that can be extremely critical for the end user, which requires freshness but also accuracy of information. However, this is often regarded as some external factors cause the information to drift towards an incorrect value [15].

The perspective of our approach is different, since it is the sender that appears as an initiator of incorrect content, due to wrong classifications in the ML unit [16]. In turn, this implies that it may be not always convenient to schedule an extra transmission, which is harmless in the case of perfect data classification, but can lead to increasing the penalty especially if the probability of misclassification is high [23].

III. Analysis

We consider a device located in a constrained environment and equipped with a pretrained ML model to classify data collected through its sensors. Because the ML classification is not perfect [16], there is a non-zero probability of misclassification p. When a data sample is misclassified, applications that operate downstream would trust the classification performed,

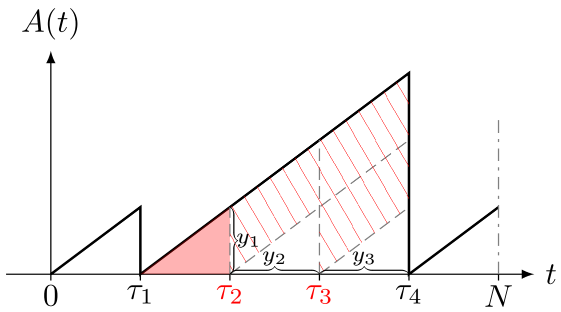

Fig. 1. AoI and AoII in the case of misclassified updates.

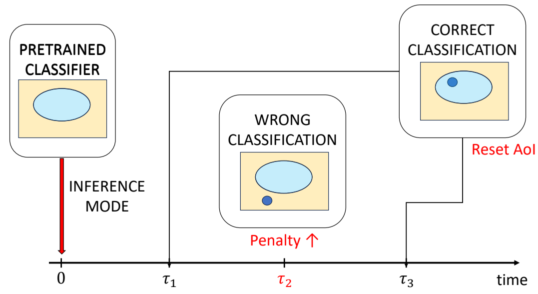

Fig. 2. Operation of the ML model. A successful classification resets AoI, otherwise the penalty due to a misclassification increases.

thinking of having fresh *and correct* information, which is prone to wrong decision making.

For this reason, we do not use just AoI as the metric of interest to be minimized, but we combine it with AoII [3], with the latter having the benefit of penalizing updates giving an incorrect reporting about the process. Fig. 1 shows the time evolution of AoII $A(t)$ in a time frame of length N with schedule update instants $\tau_1, \dots \tau_4$. If there was no misclassification, $A(t)$ would be reset to the original value, in contrast, the metric continues to increase in the case of incorrect information received at time instants τ_2 and τ_3, highlighted in red in the figure. We are interested in the average AoI, which is calculated analytically, given the process $A(t)$ as

$$\Delta = \frac{1}{N}\int_0^N A(t)dt\,. \tag{1}$$

This means that from a geometric perspective, the area below the triangle continues to increase by adding parallelograms on top of the baseline of update triangles at every update instant. Thus, we study the effect of adding a weighting parameter to these added areas with respect to the offline choice of update intervals.

We define M as the total number of updates we can perform in a unit time interval. We further define the duration of the inter-update times as $y_i = \tau_{i+1} - \tau_i$ with $\tau_0 = 0$ and $\tau_{M+1} = 1$. We consider p to be the probability that a sample is incorrectly classified and Z to be the factor that controls the importance given to errors made by the model. With these definitions and with a procedure similar to [20], the average penalty can be computed using geometric intuition as

$$\mathcal{P}(\boldsymbol{y}) = \sum_{i=0}^{M}\left[\frac{y_i^2}{2} + Z\sum_{j=i+1}^{M} y_i y_j p^{j-i}\right]. \tag{2}$$

We refer to the inner summation as ω, which is the extra component of the penalty added whenever a misclassification occurs, due to increasing AoII. We also note that for $Z=1$, (2) boils down to the standard definition of average AoI [22].

We schedule inter-update times to minimize this penalty with a stateless procedure [20]. More sophisticated methodologies can be utilized, especially leveraging the knowledge about the classification outcome in a stateful procedure [15], [29]. Yet, this approach would require a more sophisticated analysis and is therefore left for future research, also because it would require a discussion on how to detect and possibly counteract classification mistakes at run-time.

Finding a pre-determined schedule for status updates is mathematically equivalent to solving the following optimization problem

$$\min_{\boldsymbol{y}} \; \mathcal{P}(\boldsymbol{y}) \tag{3}$$

$$\text{s.t.} \quad \sum_{i=0}^{M} y_i = 1$$

$$y_i \geq 0 \quad \forall i = 0 \dots M. \tag{4}$$

In order to to find the minimum of the objective function we can null its gradient $\nabla\Delta(\boldsymbol{y})$ obtaining a system of $M-1$ linearly independent equations, which is fully determined because of the constraints.

Consisting of quadratic equations, the problem admits a solution in closed form. However, the resulting expression is cumbersome and does not add much to the discussion. Instead, the resulting optimization can be solved by numerical methods with a fairly tractable complexity. In the following, we will present results obtained by numerical optimization through the interior point method that in this specific problem allows very fast convergence [30], where solution is obtained almost immediately and with high accuracy even on personal computer processors.

IV. Machine Learning Model

A schematic of the operation procedure of the ML model, and its impact on the penalty function, is shown in Fig. 2. In more detail, we trained a support vector machine (SVM) with a radial basis function (RBF) kernel on the TUANDROMD dataset [31]. This procedure is carried out in two steps: first, we choose the best fitting model and its hyperparameters through a Bayesian optimization (BO) process [32], then we train the obtained ML model until convergence. We adopt a k-fold cross validation approach during the model selection phase.

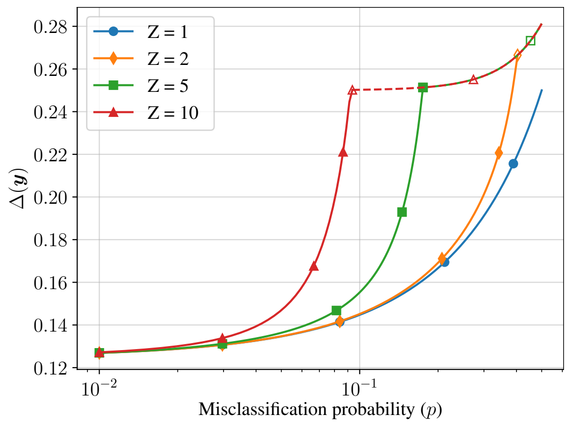

Fig. 3. Average Age of Information with 3 update instants. Dashed lines imply that multiple update instants collapse

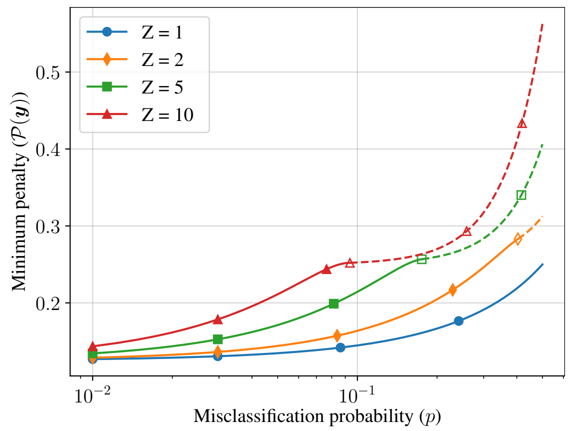

Fig. 4. Minimal penalty with 3 update instants. Dashed lines imply that multiple update instants collapse.

Both the BO procedure and the final training aim at maximizing the 1-accuracy score a of the model because this particular metric will then be used to derive the probability p of misclassification as $p = 1 - a$. Ultimately, our experiments obtained $p \approx 0.01$.

Following the initial training phase, after which the model is no longer updated and only used in inference mode, a stateless scheduler decides the best update instants a priori, based on the solution of (3). At those predefined instants τ_i, the sensing unit sample the environment and collect measurements, and the resulting data point is classified by the ML model. This determines the resulting trend of the penalty function, depending on the correctness of the classification.

We assume that all classification attempts have statistically independent outcomes, thus leaving the statistics of failures as Bernoulli with probability p and independent over different time instants. If the classification is correct, the system AoI will correspondingly be reset to 0, otherwise the penalty due to misclassification will increase. This is true even for the case when $Z = 1$, which implies that the penalty is just AoI [22], with no superimposed effect of misclassification, but becomes more acute when higher values of the coefficient Z are considering, implying a higher semantic relevance of the information content [23].

Depending on the accuracy achieved in the test set and the penalty applied to the misclassifications, we can obtain an average estimate of AoII using the specific ML model and parameters. This information is useful if we want to compare various ML models and dynamically choose which one is best to use, depending on factors such as energy consumption or latency required by the decision process [26].

In this specific contribution, we focus on the performance of a single exemplary ML model and dataset, because we are interested in analyzing the impact of a weighting factor applied to AoI in case of misdetection events [16].

V. Performance Evaluation

We hereby present the quantitative results of our evaluation, where we consider different parametric choices of p and Z and evaluate their impact on the resulting stateless scheduling, the average AoI, and the penalty defined as per (2).

In Figs. 3–6, the plots start with solid lines with full markers for lower misclassification probabilities, subsequently they switch to dashed lines with empty markers whenever the optimization dictates that multiple update instants collapse to the same value, which implies that not all the status update transmission opportunities are used. This signifies that the actual number of performed updates is lower than M, which happens for higher values of p.

As a result, AoI values tend to saturate early and the penalty experiences sudden spikes. Arguably, for higher penalty values Z the threshold after which multiple updates collapse into the same one moves towards smaller misclassification probabilities; moreover, the curves obtained have a strictly positive slope, underlying the fact that using fewer updates has a detrimental effect on the metrics of interest. Another interesting remark to be done on all the results is that since the penalty coefficient $Z = 1$ has no effect in changing the weights of the misclassification, the average AoI and the minimal penalty have the same values in the graphs for the same number of updates.

Fig. 3 describes the evolution of AoI applying the optimal scheduling with different penalty coefficients. Even though the minimization goal is the penalty function and not AoI, the latter being just one of the terms inside the penalty, it is evident that the choice of Z has a strong impact even on AoI alone. An interesting point is that when multiple updates are scheduled to be performed at the same time the expected AoI is independent of Z as all the dashed lines converge towards the same direction.

Fig. 4 shows the minimal extra penalty value paid due to misclassifications. When comparing these values to AoI alone, without extra terms due to AoII,

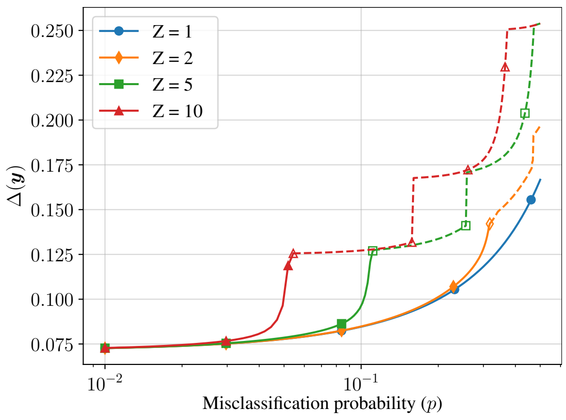

Fig. 5. Average AoI with 6 update instants. Dashed lines imply that multiple update instants collapse.

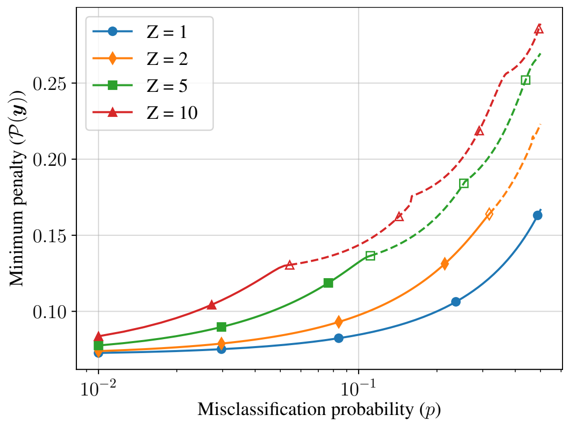

Fig. 6. Minimal penalty with 6 update instants possible. Dashed lines imply that multiple update instants collapse. Further step increases imply more updates collapsing together.

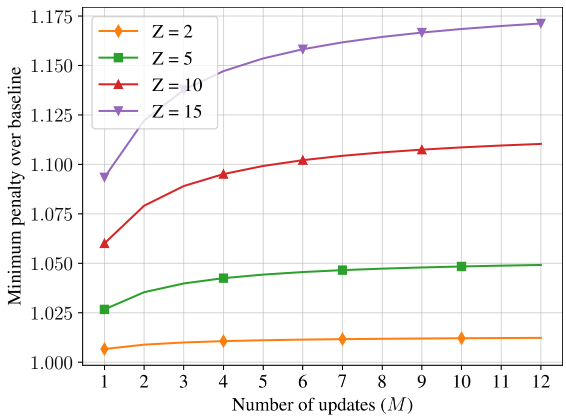

Fig. 7. Increase of the penalty for different values of the Z parameter with respect to $Z = 1$.

the increase is immediately noticeable, hence implying that misclassifications may have a severe impact on the resulting performance. When there are overlaps on the updates instants, the dashed lines do not converge to the same general direction for all the penalty coefficients. This was expected, as we are plotting the weighted area of the added parallelograms ω and thus the weight factor plays an important role in this computation.

Figs. 5 and 6 display the expected AoI and the minimal penalty $\mathcal{P}(\boldsymbol{y})$ for up to $M = 6$ distinct update instants. In this situation, the step-like behavior of the curves is enhanced by the fact that there are multiple opportunities for the scheduling of the updates instants to coincide. Moreover, the values obtained for AoI alone and the penalty due to misclassifications are very similar.

As stated for Fig. 3, the dashed lines for AoI tend to converge towards the same directions when the same number of updates is used. Comparing Fig. 6 with Fig. 4 highlights the higher degree of freedom from a larger number of possible update instants in reducing the impact of possible errors made by the ML model.

In Fig. 7, we report the increase of the penalty for different numbers of update instants M and values of the weighting parameter Z, with respect to the baseline case $Z = 1$ for the ML model described in Section IV. For $M = 1, \ldots, 5$, there is a considerable increase of the penalty that eventually wears out for a higher number of update instants available. An interesting point is that even for very high penalty factors, the worsening of the experienced penalty is below 20%, which can be attributed to the optimality of the scheduling process keeping under control surges in the penalty.

VI. Conclusion

Our study explored how misclassification in ML affects freshness and accuracy in sensing systems exchanging status updates, leading to a tradeoff [19]. We derived the expression of a penalty function, combining the expected values of ages of correct and incorrect information for a stateless scheduling [20], and we showed that finding the proper balance in the aforementioned tradeoff sometimes imply to discard some of the updates.

In addition to the analytical framework, we also performed evaluations via simulation for a real ML model [24], demonstrating that even with higher weighting factors, the AoI penalty remains bounded in the optimal scheduling. This suggests possible further studies to understand the practicality and robustness of ML in real-time status updates, so as to obtain guidance for system design and optimization.

References

[1] A. Zancanaro, G. Cisotto, and L. Badia, "Tackling age of information in access policies for sensing ecosystems," *Sensors*, vol. 23, no. 7, p. 3456, 2023.

[2] A. M. Bedewy, Y. Sun, R. Singh, and N. B. Shroff, "Low-power status updates via sleep-wake scheduling," *IEEE/ACM Trans. Netw.*, vol. 29, no. 5, pp. 2129–2141, Oct. 2021.

[3] A. Maatouk, S. Kriouile, M. Assaad, and A. Ephremides, "The age of incorrect information: A new performance metric for status updates," *IEEE/ACM Trans. Netw.*, vol. 28, no. 5, pp. 2215–2228, May 2020.

[4] M. Bakirci and M. Cetin, "Utilization of a vehicle's on-board diagnostics to reduce GPS-sourced positioning error," in *Proc. IEEE ASYU*, 2022.

[5] M. Ibrahim, A. Rostami, B. Yu, H. Liu, M. Jawahar, V. Nguyen, M. Gruteser, F. Bai, and R. Howard, "Wi-Go: accurate and scalable vehicle positioning using WiFi fine timing measurement," in *Proc. ACM MobiSys*, 2020, pp. 312–324.

[6] U. Michieli and L. Badia, "Game theoretic analysis of road user safety scenarios involving autonomous vehicles," in *Proc. IEEE PIMRC*, 2018, pp. 1377–1381.

[7] N. Zahradka, S. Geoghan, H. Watson, E. Goldberg, A. Wolfberg, and M. Wilkes, "Assessment of remote vital sign monitoring and alarms in a real-world healthcare at home dataset," *Bioengineering*, vol. 10, no. 1, p. 37, 2022.

[8] G. Cisotto, A. V. Guglielmi, L. Badia, and A. Zanella, "Joint compression of EEG and EMG signals for wireless biometrics," in *Proc. IEEE Globecom*, 2018.

[9] G. A. Adam, C.-H. K. Chang, B. Haibe-Kains, and A. Goldenberg, "Error amplification when updating deployed machine learning models," in *Proc. MLHC*, 2022, pp. 715–740.

[10] R. D. Yates, Y. Sun, D. R. Brown, S. K. Kaul, E. Modiano, and S. Ulukus, "Age of information: An introduction and survey," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 5, pp. 1183–1210, May 2021.

[11] I. Kadota, A. Sinha, E. Uysal-Biyikoglu, R. Singh, and E. Modiano, "Scheduling policies for minimizing age of information in broadcast wireless networks," *IEEE/ACM Trans. Netw.*, vol. 26, no. 6, pp. 2637–2650, 2018.

[12] Y. Sun, E. Uysal-Biyikoglu, R. D. Yates, C. E. Koksal, and N. B. Shroff, "Update or wait: How to keep your data fresh," *IEEE Trans. Inf. Theory*, vol. 63, no. 11, pp. 7492–7508, Nov. 2017.

[13] A. Buratto and L. Badia, "Analysis of age of information in slotted ALOHA networks with different strategic backoff schemes," in *Proc. IEEE CAMAD*, 2023.

[14] L. Badia, "On the impact of correlated arrivals and errors on ARQ delay terms," *IEEE Trans. Commun.*, vol. 57, no. 2, pp. 334–338, 2009.

[15] Y. Chen and A. Ephremides, "Scheduling to minimize age of incorrect information with imperfect channel state information," *Entropy*, vol. 23, no. 12, p. 1572, 2021.

[16] M. Sensoy, M. Saleki, S. Julier, R. Aydogan, and J. Reid, "Misclassification risk and uncertainty quantification in deep classifiers," in *Proc. IEEE/CVF WACV*, 2021, pp. 2484–2492.

[17] R. Hazra, M. Banerjee, and L. Badia, "Machine learning for breast cancer classification with ANN and decision tree," in *Proc. IEEE IEMCON*, 2020, pp. 0522–0527.

[18] C. Kam, S. Kompella, and A. Ephremides, "Age of incorrect information for remote estimation of a binary Markov source," in *Proc. IEEE Infocom Wkshps*, 2020.

[19] O. Ayan, M. Vilgelm, M. Klügel, S. Hirche, and W. Kellerer, "Age-of-information vs. value-of-information scheduling for cellular networked control systems," in *Proc. ACM/IEEE ICCPS*, 2019, pp. 109–117.

[20] A. Munari and L. Badia, "The role of feedback in AoI optimization under limited transmission opportunities," in *Proc. IEEE Globecom*, 2022.

[21] L. Badia, "Analysis of age of information under SR ARQ," *IEEE Commun. Lett.*, vol. 27, no. 9, pp. 2308–2312, Sep. 2023.

[22] A. Javani, M. Zorgui, and Z. Wang, "On the age of information in erasure channels with feedback," in *Proc. IEEE ICC*, 2020.

[23] J. Chen, J. Wang, C. Jiang, and J. Wang, "Age of incorrect information in semantic communications for NOMA aided XR applications," *IEEE J. Sel. Topics Signal Process.*, 2023, to be published.

[24] J. Wang, Q. Chen, and Y. Chen, "RBF kernel based support vector machine with universal approximation and its application," in *Proc. Int. Symp. Neur. Netw.* Springer, 2004, pp. 512–517.

[25] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *Proc. IEEE Infocom*, 2012, pp. 2731–2735.

[26] A. Buratto, A. Mora, A. Bujari, and L. Badia, "Game theoretic analysis of AoI efficiency for participatory and federated data ecosystems," in *Proc. IEEE ICC Wkshps*, 2023.

[27] R. D. Yates and S. K. Kaul, "The age of information: Real-time status updating by multiple sources," *IEEE Trans. Inf. Theory*, vol. 65, no. 3, pp. 1807–1827, Mar. 2018.

[28] R. Talak, S. Karaman, and E. Modiano, "Distributed scheduling algorithms for optimizing information freshness in wireless networks," in *Proc. IEEE SPAWC*, 2018.

[29] Y.-P. Hsu, E. Modiano, and L. Duan, "Scheduling algorithms for minimizing age of information in wireless broadcast networks with random arrivals," *IEEE Trans. Mobile Comput.*, vol. 19, no. 12, pp. 2903–2915, 2019.

[30] I. J. Lustig, R. E. Marsten, and D. F. Shanno, "Interior point methods for linear programming: Computational state of the art," *ORSA J. Comput.*, vol. 6, no. 1, pp. 1–14, 1994.

[31] P. Borah and D. K. Bhattacharyya, "TUANDROMD (Tezpur university android malware dataset)," UCI Machine Learning Repository, 2023.

[32] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas, "Taking the human out of the loop: A review of Bayesian optimization," *Proc. IEEE*, vol. 104, no. 1, pp. 148–175, 2016.