engineering proceedings

MDPI

Proceeding Paper

On the Clinical Use of Artificial Intelligence and Hematological Measurements for a Rapid Diagnosis and Care of Pediatric Malaria Patients in West Africa †

Ejay Nsugbe [1,*], Dephney Mathebula [2], Evi Viza [3], Oluwarotimi W. Samuel [4], Stephanie Connelly [5] and Ian Mutanga [6]

[1] Nsugbe Research Labs
[2] University of South Africa; mathed2@unisa.ac.za
[3] University of West of Scotland; evi.viza@uws.ac.uk
[4] School of Computing, University of Derby, Derby, UK; o.samuel@derby.ac.uk
[5] Hereford County Hospital, Wye Valley NHS Trust, UK; stephconnelly12@gmail.com
[6] University of Bristol; ianrodney008@gmail.com
* Correspondence: ennsugbe@yahoo.com
† Presented at the 10th International Electronic Conference on Sensors and Applications (ECSA-10), 15–30 November 2023; Available online: https://ecsa-10.sciforum.net/.

Abstract: Malaria continues to be a major cause of death worldwide, with a broad range of people spread over 90 countries being at risk of contracting the disease, and a significant cause of death in children under the age of 5. Due to this, there continues to be substantial investment towards not just the treatment of the disease, but also a more rapid and accurate means towards its diagnosis. In this work, we look to explore how measurements obtained from the complete blood count (CBC) technique from patients' blood, alongside artificial intelligence (AI) methods, could form an affordable analytical pipeline that could be adopted in hospital settings in both developed and developing countries. As part of this work, we utilize patient blood measurements acquired from pediatric patients from Ghana, West Africa, alongside various configurations of AI models towards distinguishing between malaria vs non-malaria cases in a sample set comprising over 2000 patients. Class balancing algorithms are utilized to first balance the classes for the various patient groups, followed by the use of AI algorithms to train machine learning models to differentiate between a malaria vs a non-malaria patient. The results showcased a generally high prediction accuracy, especially in the case of models with nonlinear decision boundaries, therein showing how the proposed analytic pipeline can serve as a high-throughput approach towards tackling the malaria epidemic from a diagnostics perspective and ultimately enhancing patient care strategies.

Keywords: malaria; machine learning; artificial intelligence; epidemiology; pediatrics; hematology; West-Africa; diagnosis; decision support

Citation: Nsugbe, E.; Mathebula, D.; Viza, E.; Samuel, O.W.; Connelly, S.; Mutanga, I. On the Clinical Use of Artificial Intelligence and Hematological Measurements for a Rapid Diagnosis and Care of Pediatric Malaria Patients in West Africa. *Eng. Proc.* **2023**, *56*, x. https://doi.org/10.3390/xxxxx

Academic Editor(s): Name

Published: 15 November 2023

1. Introduction

Malaria is a widespread disease which is caused by a number of "source" parasites implanted by mosquitos, and manifests itself in human beings with an array of symptoms [1]. Although malaria occurs globally, there are certain hotspots where malaria tends to occur more frequently, including sub-Saharan Africa and South East Asia [1]. In terms of health statistics, the World Health Organization (WHO) approximated that in the year 2020, around 241 million cases of malaria were reported, with 627,000 of these resulting in death, and the majority being children in sub-Saharan Africa [1].

The full impact of malaria transcends the death toll since it is also a source of financial burden due to the costs of clinical care for the affected economies [1]. Moreover, this effect

is amplified by the fact that the bulk of these economies are "developing" nations, therein contributing towards the cycle of disease and poverty in these regions [1].

Malaria parasites are only transmitted through bites of female mosquitos of the genus *Anopheles* [1]. The transmission mode first involves the mosquito biting a person who is infected, and then subsequently passing on parasites with its saliva to the next human being when it feeds on their blood [1]. Figure 1 illustrates the malaria infection cycle.

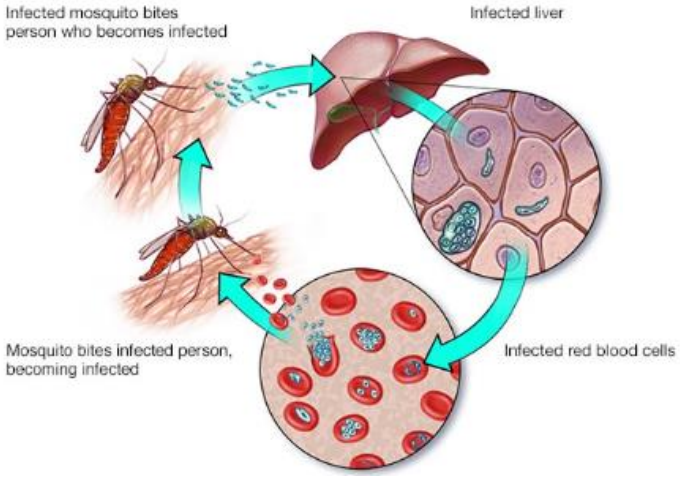

Figure 1. The malaria infection cycle [2].

Plasmodium falciparum is viewed as the most lethal of the malaria parasites, with a large number of casualties in sub-Saharan Africa in particular, where young children under the age of 5 as well as pregnant women are seen to be the most susceptible cohort [1]. Due to the nature of the disease, the WHO has invested in the order of billions towards the diagnosis, control and elimination of the global malaria issue, with a substantial amount being devoted towards treatment therapies [3]. Despite this, it has been seen that the effectiveness of the treatment continues to be reliant on the accuracy of the initial diagnosis, where progressive use of an incorrect treatment could lead to wrongful drug resistance and even death in a handful of cases [3–5]. The generalized standard for the diagnosis includes the likes of the electron microscopy method, while recently rapid diagnosis tests (RDTs) have gained a considerable amount of traction due to their ability to be used as a point of care tool [6]. However, the drawbacks of RDTs include poor detection sensitivity when parasitemia is low [6]. All of this has borne a form of motivation for an improved means towards the diagnosis of the presence of malaria parasites. The complete blood count (CBC) method has become popular in both developed and developing environments, whereby a key correlative factor of the presence of severe malaria includes a decreased count of platelets [7].

Previous related work in this area has sought to combine results from lab-based measurements alongside machine learning methods towards an automated and potential high-throughput mechanism for the diagnosis of malaria [6]. Artificial intelligence and machine learning itself have become staple mechanisms in the area of clinical medicine, with a particular appeal of providing a data-driven pattern recognition in a variety of cases, spanning areas such as oncology, anesthesia, pregnancy medicine, psychiatry, and rehabilitation, to name a few [8–13]. With respect to this particular work, related literature applied machine learning in various capacities towards the prediction of malaria, one of which used a sample size of 376 patients for the model build exercises [6,14–17]. This has been critiqued for having an overly concise sample size to converge on meaningful conclusions within the area [6]. Advancement within the research area was achieved by Morang'a et al., who managed to assemble data from a cumulative amount of 2207 patients from a number of regions in Ghana, for which various degrees of malaria were

classified as part of their study using purely hematological parameters [6]. The main shortcoming of this study was that their machine learning models were tuned and optimized using what has been perceived as an unbalanced dataset, which would lead to prediction biases in all inference and prediction actions. With the intention of advancing the body of knowledge in this area, we aim to apply class balancing algorithms towards first balancing the various data classes as a means towards negating the effect of the model, followed by the prediction of whether or not a patient has the malaria illness, and finally a differentiation between the malaria types, i.e., uncomplicated malaria (UM), non-malaria infections (nMI) and severe malaria (SM).

As a first step towards this, we utilize the broad comprehensive malaria dataset from Morang'a et al. of 2207 patients spread across Accra, Kintampo and Navrongo in Ghana, West Africa [6]. As part of this paper, a combination of the CBC measurements and machine learning models are utilized towards first reaching an overall prediction on whether or not a patient has the malaria parasite.

2. Dataset and Methods

Dataset

The dataset used in this research was acquired according to the reporting of diagnostic accuracy studies (STARD) guidelines, by Morang'a et al., where the patient inclusion criteria were as follows [6]:

SM: within the ages of 6–59 months with fever within the last 24 h and admitted to the Navrongo War Memorial Hospital, residence within the Navrongo Health and Demographic Surveillance System area, and informed consent [6]. SM is defined as hemoglobin standards of <5 g/dL, or hematocrit <15%, for which ethical approval was granted by the Noguchi Memorial Institute of Medical Research (NMIMR) review board, Naval Medical Research Center review board, and Ghana Health Service Ethics Review Committee (GHS-ERC) [6].

nMI and UM: patients for this cohort were recruited across a number of hospitals spanning Kintampo North-Municipal Hospital, Kintampo, and Ledzokuku Krowor Municipal Assembly Hospital (LEKMA), Accra [6]. The inclusion criteria was as follows: in the age range of 1–15 years, fever within the last 24 h and a signed informed consent, with ethical approval also obtained for the study [6]. A case of nMI is defined as a malaria case that is identified as negative, via microscopic validation.

CBC Measurements: are based around the use of hematological analysis of the cells within blood, which perform a characterization based on population density and morphology, from which the concentration of hemoglobin is estimated from the measurements of the red blood cells [7,18]. Various scholars have provided incremental contributions towards the effectiveness of the technology, which culminated nicely with the Coulter principle, which utilizes electrical impedance properties to perform a blood count purpose and continues to be a method of appeal and application to this very day, although optical measurements methods have now also been steadily introduced [19].

Sample Collection: as part of the sample collection process, venous blood was collected with the tourniquet not applied beyond a single minute in order to minimize erroneous measurements [6]. The sample acquisition occurred between 8 a.m. and 12 p.m. daily in order to minimize the influence of external factors such as rehydration and ingestion of food [6]. All samples were typically analyzed immediately, and those that were not analyzed within 2 h of acquisition were stored in a chamber of 2–8 degrees Celsius in order to stall any hematological changes in the sample sets [6]. It should be noted that no capillary blood samples were acquired as part of the study as they represent variations from venous blood parameters [6]. The CBC analytics was done using the automated ABX Micros 60 Hematology Analyzer, which characterizes both white and red blood cell parameters, in addition to platelets, while all data were cross-referenced as a means of validation in order to ensure consistency through the collection procedure [6].

Features and Machine Learning: the following CBC features were used as part of the analytics done in this paper: white blood cell counts, red blood cell counts, hemoglobin level, hematocrit, mean cell volume, mean corpuscular hemoglobin, mean cell hemoglobin concentration, platelet count, platelet distribution width, mean platelet volume, neutrophils percent, lymphocytes percent, mixed cells percent, neutrophils count, lymphocytes count, mixed cells count, red blood cells distribution width percent, fever symptom, and temperature [6].

The microscopy readings were utilized as the labels for the data, i.e., malaria vs no malaria, while the SMOTE algorithm was applied towards serving as class balancing purposes, which resulted in a total number of 2304 samples being utilized for all the subsequent analytics [20].

Machine Learning: the following machine learning models were used as part of the pattern recognition exercises: decision tree (DT), linear discriminant analysis (LDA), logistic regression (LR), support vector machines (SVM), linear-SVM (LSVM), quadratic-SVM (QSVM), cubic-SVM (CSVM), fine Gaussian-SVM (FGSVM), k-nearest neighbors (KNN) [8]. All classification models were validated with the K-fold validation scheme, with K selected as 10.

3. Results

The results in Table 1 show the model-based prediction performance for the classification of malaria vs non-malaria from the aforementioned CBC measurements. From the results Table 1, it can be seen that the prediction results are generally high across all the various candidate classification models, of which the QSVM and CSVM (characterized by the nonlinear Kernels) both achieved the best classification performance of 89% accuracy, therein showing that a nonlinear decision boundary is optimal for the separation of the two data classes. Besides, it can be seen that the fine Gaussian-SVM and the Decision Tree classifiers yielded the least classification performance (84% and 85%) in comparison to the other classification methods. These results reflect a more robust model performance when compared with that of Morang'a et al. whose model build and design was done with the use of an unbalanced sample set, which ultimately would probably result in a biased model prediction.

Table 1. Model prediction performance for classification of malaria vs non-malaria from CBC measurements.

Model	Accuracy (%)
DT	85
LDA	86
LR	86
LSVM	87
QSVM	89
CSVM	89
FGSVM	84
KNN	87

The achievement of this would potentially allow for a deployment of a high throughput model pipeline which would allow for a quicker diagnosis of malaria patients from their blood samples, CBC analysis and model-based predictions. A pipeline of this can be

adopted and deployed in both developed and developing economies, due to the nature of the tools and model adopted. An illustration of the proposed flow can be seen in Figure 2.

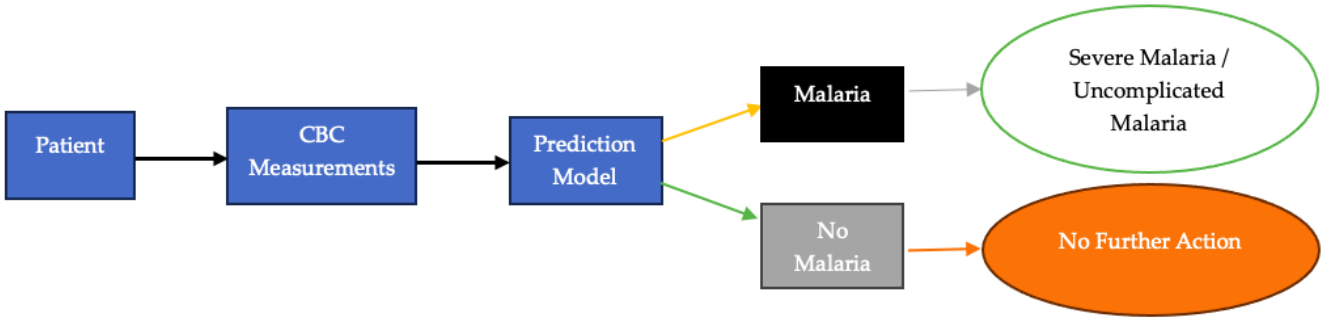

Figure 2. Pipeline for quicker diagnosis of malaria from blood samples, CBC analysis and model-based predictions.

The pipeline diagram shows a multi-stage model comprising the various sub-processes needed to come together in an ensemble-like fashion to produce the intended outcome of a high-throughput malaria prediction system. This features the lab-based CBC measurements which are used towards the training and optimising of the machine learning model with the intended role to carry out a tier-like prediction scheme. This includes, first, a surface prediction of whether or not the malaria parasite exists within the sample of the patient (focus of this paper), followed by a rating of the class of malaria where possible, i.e., SM or UM, which would provide key information that would assist in the prioritisation of patient care (area of focus for further work).

A visualisation of the data using the principal component analysis (PCA) can be seen in Figure 3, where it can be observed that there exists a degree of overlap between the various data classes; hence the optimal decision boundary appeared to be of a nonlinear nature due to the apparent overlap between the classes. The inclusion of further features within the feature vector would enhance the class separatability of the various data classes, a notion which would be explored deeper in subsequent work on this topic.

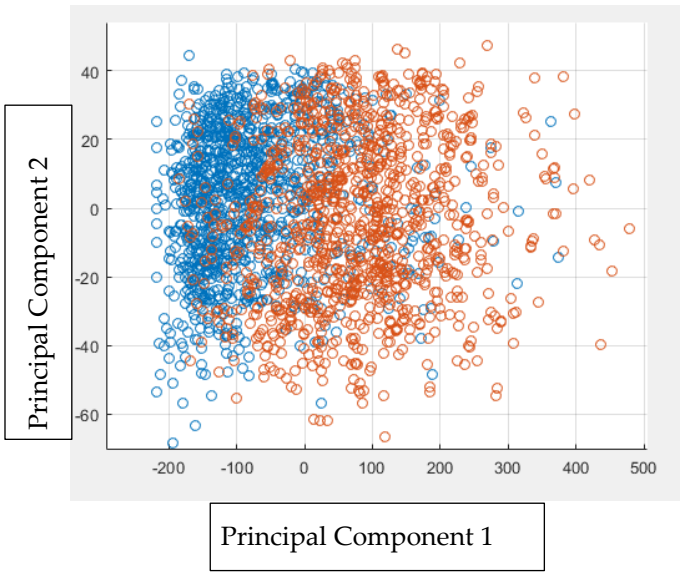

Figure 3. Principal component analysis of the data (Where Blue-Malaria and Red-Non-Malaria).

The computational complexity of machine learning models is a pivotal consideration in their practical deployment. Briefly, it encompasses the amount of computational

resources, time, and memory required to train and deploy these models. The computational complexity of a model depends on various factors, including the size of the dataset, the number of features, the model's architecture and parameters of the algorithm. While we have focused on classification accuracy as a metric thus far, it is worth noting that the computational complexity of the models should be further investigated.

4. Conclusions

Malaria is a global disease that has been identified by the WHO as an epidemic, with a large portion of related deaths centering around children in sub-Saharan Africa. The negative effects of malaria within these countries are profound in terms of healthcare and economics. The CBC represents a measurement technique largely employed in the diagnosis of malaria, which has been adopted in both developed and developing regions. Prior work has adopted the use of artificial intelligence as a means towards pattern recognizing measurements which correlate towards a subsequent malaria diagnosis. The shortcoming of this involved the use of an unbalanced dataset, which would lead to biased predictions by the concurrent prediction machine. The contributions made in this paper involve the use of class balancing algorithms, firstly to the pediatric malaria dataset, alongside the formulation of a multistage pipeline which looks to provide an initial binary diagnosis of whether a patient has the malaria parasite, followed by an associated prediction of the extent of the malaria.

The obtained results showed that the various adopted models have good prowess in the prediction of whether a patient has malaria, where the SVM with a quadratic nonlinear decision boundary was seen to be optimal for differentiating between the two classes of pediatric patients. As a means towards improving this, subsequent work in this area would now involve the potential inclusion of further features within the feature vector to potentially induce class separability with a view towards a higher-class differentiation result. This would be followed by addressing the second portion of the prediction model, aimed around the identification of the extent of the malaria, given a positive malaria diagnosis. Further work in this area would also involve the potential projection of the features in a time-series format, which would enable the use of classical signal processing and signal decomposition toolsets as applied in previous studies [21–23].

Author Contributions: All authors contributed equally to this manuscript. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: The study was conducted within ethical standards, details of which have been included within the manuscript

Informed Consent Statement: Written consent was obtained from all patients

Data Availability Statement: The data used is available via a cited repository within the manuscript

Acknowledgments: The authors would like to thank Brian Kerr from Kerr Editing for proofreading this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. CDC-Centers for Disease Control and Prevention CDC-Malaria-FAQs. Available online: https://www.cdc.gov/malaria/about/faqs.html (accessed on 6 July 2023).
2. Malaria Transmission Cycle. Available online: https://www.mayoclinic.org/diseases-conditions/malaria/multimedia/malaria-transmission-cycle/img-20006373 (accessed on 6 July 2023).
3. World Health Organization. *World Malaria Report 2018*; World Health Organization: Geneva, Switzerland, 2018; ISBN 978-92-4-156565-3.
4. Watson, O.J.; Sumner, K.M.; Janko, M.; Goel, V.; Winskill, P.; Slater, H.C.; Ghani, A.; Meshnick, S.R.; Parr, J.B. False-Negative Malaria Rapid Diagnostic Test Results and Their Impact on Community-Based Malaria Surveys in Sub-Saharan Africa. *BMJ Glob. Health* **2019**, *4*, e001582. https://doi.org/10.1136/bmjgh-2019-001582.

5. Mouatcho, J.C.; Goldring, J.P.D. Malaria Rapid Diagnostic Tests: Challenges and Prospects. *J. Med. Microbiol.* **2013**, *62*, 1491–1505. https://doi.org/10.1099/jmm.0.052506-0.
6. Morang'a, C.M.; Amenga-Etego, L.; Bah, S.Y.; Appiah, V.; Amuzu, D.S.Y.; Amoako, N.; Abugri, J.; Oduro, A.R.; Cunnington, A.J.; Awandare, G.A.; et al. Machine Learning Approaches Classify Clinical Malaria Outcomes Based on Haematological Parameters. *BMC Med.* **2020**, *18*, 375. https://doi.org/10.1186/s12916-020-01823-3.
7. Healthdirect Australia Full Blood Count (FBC)—A Common Blood Test. Available online: https://www.healthdirect.gov.au/full-blood-count (accessed on 6 July 2023).
8. Nsugbe, E.; Ser, H.-L.; Ong, H.-F.; Ming, L.C.; Goh, K.-W.; Goh, B.-H.; Lee, W.-L. On an Affordable Approach towards the Diagnosis and Care for Prostate Cancer Patients Using Urine, FTIR and Prediction Machines. *Diagnostics* **2022**, *12*, 2099. https://doi.org/10.3390/diagnostics12092099.
9. Nsugbe, E.; Connelly, S. Multiscale Depth of Anaesthesia Prediction for Surgery Using Frontal Cortex Electroencephalography. *Healthc. Technol. Lett.* **2022**, *9*, 43–53. https://doi.org/10.1049/htl2.12025.
10. Nsugbe, E.; Obajemu, O.; Samuel, O.W.; Sanusi, I. Application of Noninvasive Magnetomyography in Labour Imminency Prediction for Term and Preterm Pregnancies and Ethnicity Specific Labour Prediction. *Mach. Learn. Appl.* **2021**, *5*, 100066. https://doi.org/10.1016/j.mlwa.2021.100066.
11. Nsugbe, E. On the Application of Metaheuristics and Deep Wavelet Scattering Decompositions for the Prediction of Adolescent Psychosis Using EEG Brain Wave Signals. *Digit. Technol. Res. Appl.* **2022**, *1*, 9–24. https://doi.org/10.54963/dtra.v1i2.40.
12. Nsugbe, E.; Williams Samuel, O.; Asogbon, M.G.; Li, G. Contrast of Multi-Resolution Analysis Approach to Transhumeral Phantom Motion Decoding. *CAAI Trans. Intell. Technol.* **2021**, *6*, 360–375. https://doi.org/10.1049/cit2.12039.
13. Nsugbe, E.; Phillips, C.; Fraser, M.; McIntosh, J. Gesture Recognition for Transhumeral Prosthesis Control Using EMG and NIR. *IET Cyber-Syst. Robot.* **2020**, *2*, 122–131. https://doi.org/10.1049/iet-csr.2020.0008.
14. Poostchi, M.; Silamut, K.; Maude, R.J.; Jaeger, S.; Thoma, G. Image Analysis and Machine Learning for Detecting Malaria. *Transl. Res.* **2018**, *194*, 36–55. https://doi.org/10.1016/j.trsl.2017.12.004.
15. Bediako, Y.; Adams, R.; Reid, A.J.; Valletta, J.J.; Ndungu, F.M.; Sodenkamp, J.; Mwacharo, J.; Ngoi, J.M.; Kimani, D.; Kai, O.; et al. Repeated Clinical Malaria Episodes Are Associated with Modification of the Immune System in Children. *BMC Med.* **2019**, *17*, 60. https://doi.org/10.1186/s12916-019-1292-y.
16. KalantarMotamedi, Y.; Eastman, R.T.; Guha, R.; Bender, A. A Systematic and Prospectively Validated Approach for Identifying Synergistic Drug Combinations against Malaria. *Malar. J.* **2018**, *17*, 160. https://doi.org/10.1186/s12936-018-2294-5.
17. Shrinet, J.; Nandal, U.K.; Adak, T.; Bhatnagar, R.K.; Sunil, S. Inference of the Oxidative Stress Network in Anopheles Stephensi upon Plasmodium Infection. *PLoS ONE* **2014**, *9*, e114461. https://doi.org/10.1371/journal.pone.0114461.
18. Blood Tests for CLL. Available online: https://www.cancerresearchuk.org/about-cancer/chronic-lymphocytic-leukaemia-cll/getting-diagnosed/tests/blood-tests (accessed on 6 July 2023).
19. Graham, M.D. The Coulter Principle: Foundation of an Industry. *JALA J. Assoc. Lab. Autom.* **2003**, *8*, 72–81. https://doi.org/10.1016/S1535-5535-03-00023-6.
20. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-Sampling Technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. https://doi.org/10.1613/jair.953.
21. Nsugbe, E.; Starr, A.; Ruiz-Carcel, C. Monitoring the Particle Size Distribution of a Powder Mixing Process with Acoustic Emissions: A Review. *Eng. Technol. Ref.* **2016**, *1*, 1–12.
22. Nsugbe, E. Particle Size Distribution Estimation of a Powder Agglomeration Process Using Acoustic Emissions. Ph.D. Thesis, Cranfield University, Cranfield, UK, 2017.
23. Nsugbe, E.; Ruiz-Carcel, C.; Starr, A.; Jennions, I. Estimation of Fine and Oversize Particle Ratio in a Heterogeneous Compound with Acoustic Emissions. *Sensors* **2018**, *18*, 851. https://doi.org/10.3390/s18030851.