A deep learning algorithm for potato tuber hollow heart classification

Arash Abbasi*, Max J. Feldman†, Jaebum Park‡, Katelyn Greene†, Richard G. Novy‡, Huaping Liu§

*The Beacom College of Computer and Cyber Sciences, Dakota State University, South Dakota, USA
Arash.Abbasi@@dsu.edu

†Temperate Tree Fruit and Vegetable Research Unit, USDA – Agricultural Research Service, Prosser, WA
{Max.Feldman, Katelyn.Greene}@usda.gov

‡Small Grains and Potato Germplasm Research Unit, USDA – Agricultural Research Service, Aberdeen, ID
{Jaebum.Park, Rich.Novy}@usda.gov

§Electrical Engineering and Computer Science, Oregon State University, Oregon, USA
Huaping.Liu@oregonstate.edu

Abstract—A novel deep learning algorithm is proposed for hollow heart detection which is an internal tuber defect. Hollow heart is one of many internal defects that decrease the market value of potatoes in the fresh market and food processing sectors. Susceptibility to internal defects like the hollow heart is influenced by genetic and environmental factors so elimination of defect-prone material in potato breeding programs is important. Current methods of evaluation utilize human scoring which is limiting (only collects binary data) relative to the data collection capacity afforded by computer vision or are based upon X-ray transmission techniques that are both expensive and can be hazardous. Automation of defect classification (e.g. hollow heart) from data sets collected using inexpensive, consumer-grade hardware has the potential to increase throughput and reduce bias in public breeding programs. The proposed algorithm consists of ResNet50 as the backbone of the model followed by a shallow fully connected network (FCN). A simple augmentation technique is performed to increase the number of images in the data set. The performance of the proposed algorithm is validated by investigating metrics such as precision and the area under the curve (AUC).

Index Terms—*Solanum tuberosum, hollow heart, CNN, ResNet, deep learning, optimization*

I. INTRODUCTION

Potato (*Solanum tuberosum*) is the most economically important vegetable crop in the United States. Over the last decade, the U.S. has consistently planted over $900,000$ acres of potato, resulting in the production of over 400 million CWT of tuber material and an economic value approaching four billion U.S. dollars [1]. Roughly two-thirds of the crop is sold to the food processing industry to generate products like French fries, potato chips, dehydrated potato products, and canned goods, whereas a quarter is sold as a fresh market product to consumers in grocery stores and restaurants. Quality standards for potatoes are rigorous in both the processing and fresh market industry. Defects can occur during the growth season (hollow heart, internal browning syndrome, growth cracks, virus-induced necrotic lesions), harvest (bruise, cracking, greening), and storage (sprouting, greening); the consequence of which is a financial loss to the grower. Reducing the propensity of cultivars to express these defects is one major objective of potato breeding programs. In some cases, genetic methods can be used to predict and select against offspring that will be susceptible to defects. Unfortunately, tabulating the type and severity of tuber defects on the scale needed to evaluate early-stage breeding populations (> 200 clones; between 5 to 100 tubers per clone) is not trivial. Challenges arise as symptomology can be similar between defect type and time which can be allocated to observe any individual tuber is very small. The low-cost, high information content, and rapid acquisition time make digital imaging of tubers an efficient way to document the properties off-spring in breeding populations.

With the rapid increase in computation power and big data, deep learning has shown tremendous success in various image and signal processing applications such as health care, biology, anomaly detection [2, 3, 4, 5, 6]. Unlike conventional rule-based algorithms, in deep learning algorithms suitable discriminatory features for classification/detection are extracted by the model automatically during the training phase. Hence, better performance is achievable.

In this report, a novel deep learning approach is proposed for potato hollow heart detection. Hollow heart is a physiological, internal tuber defect that can occur when growth in the perimedullary region of the tuber outpaces growth of the pith causing the development of lens-, star-, or irregularly-shaped cavities within the pith tissue [7]. Frequently, expression of hollow heart is a consequence of water or nutrient fluctuation during the growth season [7]. Genetics has been demonstrated to play a role in hollow heart susceptibility [8] and in some cases susceptibility is likely to be independent of final tuber size [9].

The presence of the hollow heart is somewhat common and relatively easy to identify, making it an excellent test case to evaluate the feasibility of applying deep learning approaches to defect classification in plant breeding programs. A hollow heart and non-hollow heart examples are shown in Fig. 1 and Fig. 2, respectively.
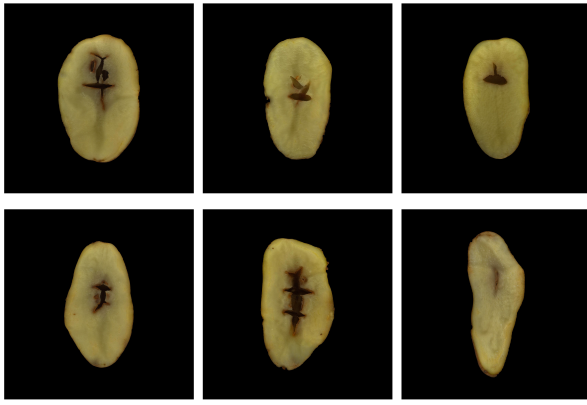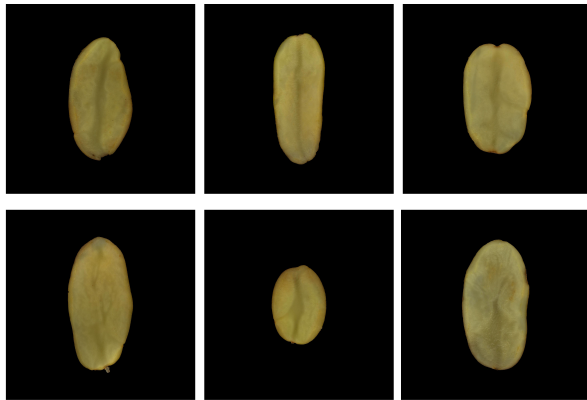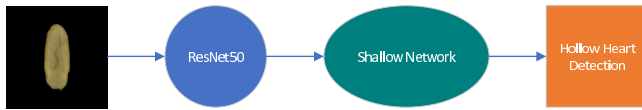
Fig. 1: Hollow heart.

Fig. 2: Non-hollow heart.

Fig. 3: A simple block diagram of the proposed model.

II. METHODS

The data set contains 1610 three-channel RGB images of potato tuber cut in half lengthwise. Each image is 1900×1900 pixels $\times$ pixels, contains a single tuber that is centered on a black background. The presence of internal tuber defects such as hollow heart, internal brown spot, bruise, anthocyanin accumulation, and tuber greening was scored manually according to a USDA visual aid guide provided in [10].

A. Plant material

The subjects are tubers derived from 189 F1 progeny from the A08241 autotetraploid linkage mapping population. In 2019, both parental lines (Palisade Russet, female; ND02873 2Russ, male) and their progeny were planted in a randomized complete block design with two replications of eight-hill plots by scientists at the USDA-ARS Small Grains and Potato Germplasm Research Unit in Aberdeen, ID.

B. Digital imaging of tubers

Computer vision measurements are acquired by scientists at the USDA-ARS Temperate Tree Fruit and Vegetable Research Unit in Prosser, WA. Digital scans of tuber internal characteristics are acquired using a Hewlett Packard HP ColorJet 6200C flatbed scanner (See Supplemental Fig. 1). Each clone was replicated twice, and each replicate contains five tubers within the image. Before scanning each tuber is cut in half lengthwise and an image of each tuber is captured. Both size and radiometric calibration standards are included in each image. A blue plastic poker chip (37 mm diameter) is used as a size standard in all images and scans. An X-Rite ColorChecker Classic is used as a radiometric standard for the top-down imaging on the black background whereas an X-Rite ColorChecker Mini is used as color standards for the scanner. Python scripts used to control the imaging hardware can be found at [11]. Individual tubers were labeled by their position in the image and manually scored for the presence or absence of internal defects according to the USDA grading standards visual aid chart.

III. PROPOSED DEEP LEARNING ARCHITECTURE

A. Optimization

The goal of the proposed algorithm is to identify hollow heart defects in the potato images. This can be sought as a binary classification task, i.e., hollow heart vs non-hollow heart classification, and hence deep learning algorithms suitable for classification applications can be employed. Let's C_0 and C_1 represent the non-hollow heart and hollow heart classes, respectively, and d an image in the data set. In this case, we aim to find the probability that the image d is a member of non-hollow heart $P(C_0|d)$ or hollow-heart $P(C_1|d)$ sets. Since there are two discrete outcome, the Bernoulli distribution function can be derived to express the conditional probability $P(y|x)$, that is the probability of the the image class is y for an observation x, as

$$P(y|x) = \hat{y}^y + (1 - \hat{y})^{1-y}, \qquad (1)$$

where $y \in \{0, 1\}$ is the true label and $\hat{y}$ is the model prediction for the image x. The log likelihood estimation can be employed that aims to maximize the $\log p(y|x)$. In this case, the log-likelihood cost-function C_{LL} is defined as

$$C_{LL} = y \log \hat{y} + (1 - y) \log(1 - \hat{y}). \qquad (2)$$

It is seen that maximizing the log-likelihood cost-function,i.e., maximizing the probability of the correct estimated labels, is equivalent to minimizing the cross-entropy between true labels and the estimated labels which is defined as

$$\begin{aligned} C_{CE} &= -\log P(y|x) \\ &= -(y \log \hat{y} + (1 - y) \log(1 - \hat{y})) \qquad (3) \\ &= C_{LL}. \end{aligned}$$

This ensures the maximization of the correct prediction and minimization of the incorrect classification. Assuming there are M independent and identically distributed (iid) images in

TABLE I: The proposed model architectures.

Model 1	ResNET50 $\rightarrow$ (Dense(256), ReLU) $\rightarrow$ (Dense(128), ReLU)$\times$ 3$\rightarrow$ (Dense(128), sigmoid) $\rightarrow$ (Dense(2), sigmoid)
Model 2	ResNET50 $\rightarrow$ (Dense(128), sigmoid) $\rightarrow$ (Dense(2), sigmoid)

the training data set, the log-likelihood cost-function for the training data set is given as

$$
\begin{aligned}
\log P(Y|X) &= \log \prod_{i=1}^{M} P(y^{(i)}|x^{(i)}) \\
&= \sum_{i=1}^{M} \log P(y^{(i)}|x^{(i)}) \\
&= \sum_{i=1}^{M} y^{(i)} \log \hat{y}^{(i)} + (1 - y^{(i)}) \log(1 - \hat{y}^{(i)}) \\
&= -\sum_{i=1}^{M} C_{CE}(y^{(i)}, \hat{y}^{(i)}),
\end{aligned}
\tag{4}
$$

where $Y = \{y_1, y_2, \cdots, y_M\}$ and $X = \{x_1, x_2, \cdots, x_M\}$ are the sets of all labels and images in the training data sets. By employing deep learning algorithms, the cross-entropy cost-function in (4) is parametrized by the weights of the model $\boldsymbol{\Theta} = (\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots, \boldsymbol{\theta}^{(L)})$, in which L represents the number of hidden layers in the model. We represent the output of the model as $f(x^{(i)}; \boldsymbol{\Theta})$ for the i-th image $x^{(i)}$. In this case, the optimum solution for the weight of the model $\boldsymbol{\Theta}$, weights that minimized the cross-entropy cost-function, can be given as

$$
\begin{aligned}
\hat{\boldsymbol{\Theta}} &= \underset{\boldsymbol{\Theta}}{\arg\min} \; C_{CE}(f(X; \boldsymbol{\Theta}), Y) \\
&= \underset{\boldsymbol{\Theta}}{\arg\min} \; \frac{1}{M} \sum_{i=1}^{M} C_{CE}\Big(f(x^{(i)}; \boldsymbol{\Theta}), y^{(i)}\Big) \\
&= \underset{\boldsymbol{\Theta}}{\arg\min} \frac{1}{M} \sum_{i=1}^{M} \bigg(y^{(i)} \log f(x^{(i)}; \boldsymbol{\Theta}) \\
&\quad + (1 - y^{(i)}) \log(1 - f(x^{(i)}; \boldsymbol{\Theta})) \bigg).
\end{aligned}
\tag{5}
$$

To minimize potential overfitting and address the sparsity in the data set, ℓ_2 and ℓ_1 regularization are employed, respectively. In this case, the cost-function can be written as

$$
\hat{\Theta} = \underset{\Theta}{\arg\min} \; C_{CE}(f(X; \Theta), Y) + \lambda \|\Theta\|_2 + \beta \|\Theta\|_1, \tag{6}
$$

where λ and α are regularization coefficients for ℓ_2 and ℓ_1 regularization, respectively. Various gradient descent algorithms can be employed that aim to find the set of In this paper, adaptive moment estimation (Adam) is used to find the set of the optimum solution for the weights of the model $\boldsymbol{\Theta}$. in (6), in terms of minimizing cross-entropy or maximizing the log-likelihood [13].

B. Proposed Deep Learning Model

In this paper, we propose two deep learning architectures that are composed of ResNet50 [12] followed by different shallow fully connected networks (FCNs). A block diagram of the proposed architectures is depicted in Figure 3. The shallow FCN is added to the ResNet50 to decrease the dimensionality of the last layer of the ResNet50 so that it will be suitable for binary classification. The proposed architectures are summarized in Table I. The cross-entropy (CE) loss derived in (4, the area under the curve, and the precision of the models are evaluated. Precision of the proposed modes is measured as

$$
precision = \frac{TP}{TP + FP}, \tag{7}
$$

where TP and FP are true positive and false positive respectively. The categorical accuracy of the model is calcualted as

$$
precision = \frac{TP + TN}{TP + FP + TN + FN}, \tag{8}
$$

where TN and FN are true negative and false negative, respectively.

IV. SIMULATION RESULTS

The data set consists of 50 potato images with the hollow heart defect and 1730 images with no hollow heart. It is obvious that the data set is unbalanced. Therefore in a prepossessing step, we aim to 1) balance the data set by selecting randomly only 50 images from the non-hollow heart defect class and 2) increase the number of the images in the data set by performing a simple augmentation procedure that is rotating the images by 90 and 180 degrees and adding additive white Gaussian noise (AWGN) with $\mu = 0$ and variance σ^2. After the prepossessing step, there are 150 potato images with hollow-heart deficiency and 150 with no hollow heart deficiency in the training data set. Therefore, in total there are 300 images in the data set.

There are $\binom{1700}{50}$ different permutations for selecting 50 images from total 1700 non-hollow heart images. To increase the model exposure to more diverse non-hollow cases, a Monte Carlo approach is employed where the model is trained 120 times independently. The dimension of the potato images is 1200×1200 pixels $\times$ pixels. The images are resized to 100×100 pixels $\times$ pixels to reduce the redundancy in the images and enhance the computational processing speed. Additionally, since the subject is plotted on a completely black background in each image and the background will not influence defect status, a cropping mechanism is employed so that only the potato portion of the images is considered in the cropped images and the background is removed from each image

The cross-entropy loss, the precision, the area under the curve (AUC), and the categorical accuracy performance of the proposed algorithms when the variance of the AWGN is $\sigma^2 = \{0.01, 0.03, 0.05\}$ and for different regularization, values
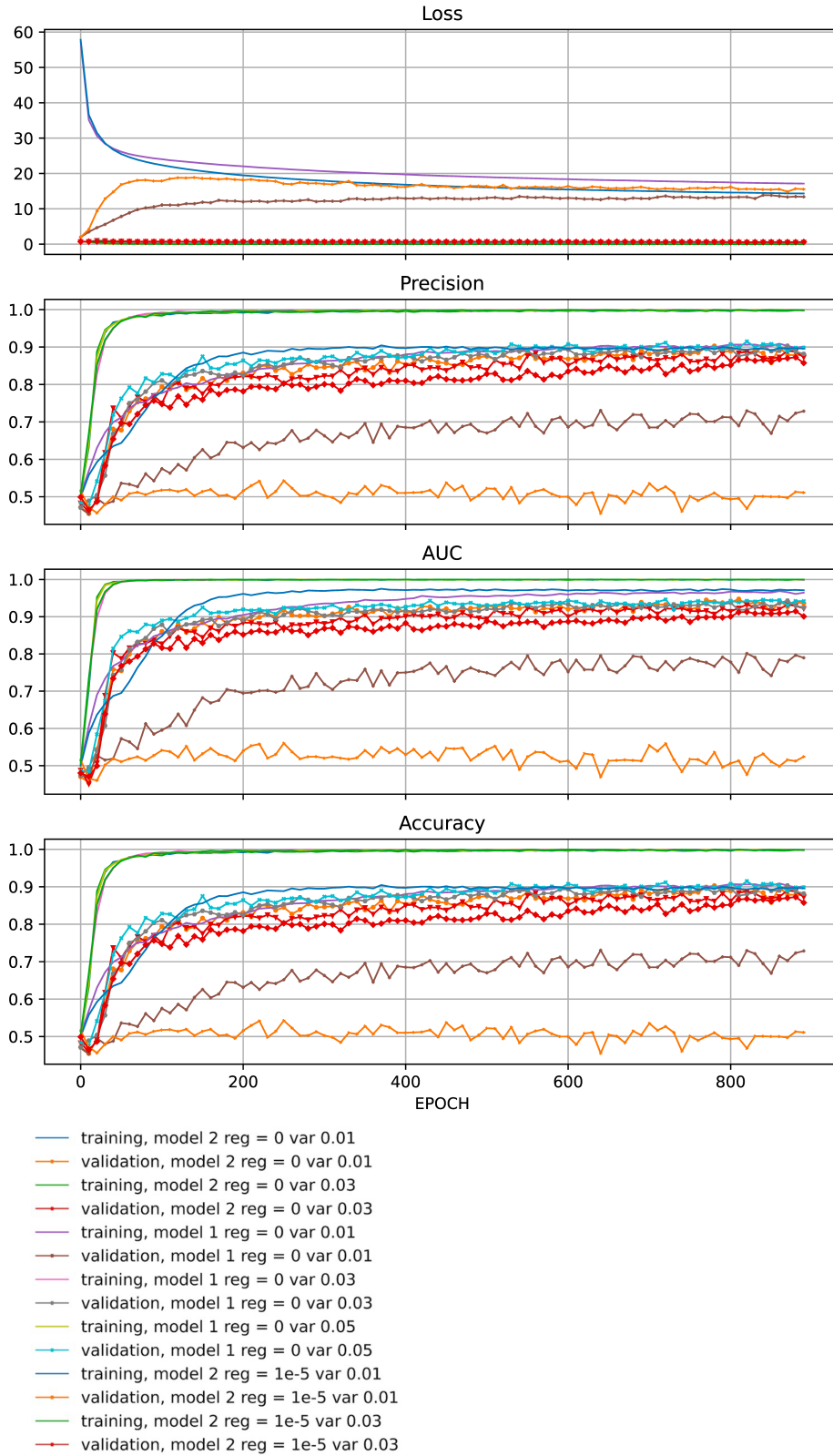
Fig. 4: The performance of the proposed models.

are depicted in Figs. 4. The learning rate, regularization parameter, and the batch size are $\{1e-3, 1e-5, 50\}$, respectively. As expected, the performance of the model in the training phase outperforms the validation performance. Moreover, it is seen that the model 1 outperforms the model 2. This demonstrates the effectiveness of the deeper FCN added to the ResNet model. In other words, as the depth of the FCN increases, the model can learn more discrepancy features from the input images and hence the performance of the model improves. Additionally, for all three metrics, the performance of the models improves as the variance of the AWGN decreases. This is because, in lower variance, the model can learn better discrepancy features from the images, and hence, better performance can be achieved. Surprisingly, however, it is seen that the performance of the models degrade when regularization is added to the models. One hypothesis for this anomaly behavior is that since the number of images in the training data set is small, hence, the added regularization terms behave restrictive for the model weights that are set to be trained. It is also seen that the categorical accuracy performance is identical to the precision. This is because the model is balanced. That is, its ability to correctly classify hollow heart (positive) samples is same as its ability to correctly classify non-hollow heart (negative) samples. The accuracy and precision performance demonstrate the capacity of the model to detect and identify the hollow heart (and therefore, non-hollow heart) cases. For example, after 800 training epochs, the model 1 can achieve 90% accuracy.

V. Discussion

This study suggests that deep learning techniques applied to image data sets collected using consumer grade hardware may be sufficient to classify the presence or absence of tuber defects in breeding populations; so long as the training set is representative of the classification problem and contains enough images to train the model. As demonstrated, augmentation of data classes using image rotation and introduction of additive white Gaussian noise may be used to increase the size of the training set but caution is advised in application settings. The model described may not perform as expected if applied to other data sets, particularly data sets containing tubers with different shapes and/or flesh colors.

Despite the need for destructive sampling, the hollow heart classification approach presented in this study offers several advantages over other commonly used diagnostic methods described in the literature. The relationship between tuber size and specific gravity has been used identify and remove potentially affected tubers in bulk grading applications [14, 15, 16]. Unfortunately, manual destructive assessment is still required to achieve the precision needed for breeding applications. Although relatively high precision has been achieved using light transmittance [17], confounding variation associated with differences in tuber diameter, tuber skin thickness, and tuber skin quality make the evaluation of the diverse tuber material encountered in breeding populations challenging. The application of non-destructive ultrasonic technologies is fraught with

difficulties due to the high attenuation coefficient of potato tubers, need for multiple transducers, and frictional noise generated by the tuber moving along the transducer [21]. Excellent success of hollow heart detection and sample throughput has been achieved using non-destructive techniques including X-ray machines [14, 15, 18, 19] and hyperspectral sensors [20]. Unfortunately, the equipment (X-ray and hyperspectral) and safety measures (X-ray) required for deployment make these techniques is cost prohibitive for many potato breeding programs. X-ray data also lacks colorimetric features which may be needed to discriminate between different types of internal defect (bruising, greening, anthocyanin accumulation, internal brown spot) present within susceptible clones of breeding populations.

By far the greatest advantages of the technique presented is the low cost of deployment and potential to develop extensible models capable of classifying additional defects from the same data set. The capacity to inexpensively perform semi-automated, high resolution data collection, and downstream defect classification within potato breeding populations will enhance our ability to understand of how defect susceptibility is inherited and improve our ability to select productive, blemish free potato cultivars.

VI. Conclusion

In this paper, a novel deep learning algorithm was proposed for potato hollow heart and non-hollow heart classification/detection. The performance of the proposed model was evaluated by investigating various metrics such as precision and the AUC. The results demonstrated the effectiveness of the model in different AWGN regions. As the next step, we aim to classify additional features (defects) from potato tubers. Many human scorable potato characteristics (sprouting, fungal/bacterial growth, bruising, and necrotic lesions) have the potential to be quantified using CNN classification algorithms. For future research, we aim to extend the proposed algorithms to detect and classify other potato defects such as internal brown spots, bruising, greening, and anthocyanin accumulation simultaneously. This requires constructing a potato data set that includes large numbers of each potato defect type. Therefore, additional data collection efforts are underway and we aim to determine if the proposed algorithms can be extended to classify these features.

Acknowledgment

References

[1] National Agricultural Statistics Service National Agricultural Statistics Service (NASS), *North American Potatoes* [Online]. Available:

https://downloads.usda.library.cornell.edu/usda-esmis/files/4m90dv511/sn00bp98b/zk51w7983/uscp1220.pdf [Accessed: Aug. 5, 2021].
Agricultural Statistics Board, United States Department of Agriculture (USDA) https://downloads.usda.library.cornell.edu/usda-esmis/files/4m90dv511/sn00bp98b/zk51w7983/uscp1220.pdf

[2] Y. -S. Lin, P. -H. Huang and Y. -Y. Chen, "Deep Learning-Based Hepatocellular Carcinoma Histopathology Image Classification: Accuracy Versus Training data set Size," *IEEE Access*, 2021, 9, pp. 33144-33157, doi: 10.1109/ACCESS.2021.3060765.

[3] S. Peng, S. Sun and Y. -D. Yao, "A Survey of Modulation Classification Using Deep Learning: Signal Representation and Data Preprocessing," *IEEE Trans. Neural Networks and Learning Systems*, doi: 10.1109/TNNLS.2021.3085433.

[4] L. Luo et al., "Deep Mining External Imperfect Data for Chest X-ray Disease Screening," *IEEE Trans. Medical Imaging*, 2020, 39, pp. 3583-3594, doi: 10.1109/TMI.2020.3000949.

[5] Z. Feng, J. Tang, Y. Dou and G. Wu, "Learning Discriminative Features for Semi-Supervised Anomaly Detection," *IEEE Int. Conf. Acoustics, Speech, and Signal Process. (ICASSP)*, 2021, pp. 2935–2939, doi: 10.1109/ICASSP39728.2021.9414285.

[6] D. Lappas, V. Argyriou and D. Makris, "Fourier Transformation Autoencoders for Anomaly Detection," *IEEE Int. Conf. Acoustics, Speech, and Signal Process. (ICASSP)*, 2021, pp. 1475–1479, doi: 10.1109/ICASSP39728.2021.9415010.

[7] B. L. Rex and G. Mazza, "Cause, control and detection of hollow heart in potatoes: A review," *American Potato Journal* 1989, 66, pp. 165-183, doi: 10.1007/BF02853679

[8] C. Zorilla, F. Navarro, S. Vega-Semorile, J. Palta "QTL for pitted scab, hollow heart, and tuber calcium identified in a tetraploid population of potato derived from an Atlantic × Superior cross," *Crop Science* 2021, 61, pp. 1630-1651, doi: 10.1002/csc2.20388

[9] S.H. Jansky and D.M. Thompson, "Expression of hollow heart in segregating tetraploid potato families," *American Journal of Potato Research* 1990, 67, pp. 695-–703, doi: 10.1007/BF03044016

[10] Agricultural Marketing Service - USDA Official Visual Aids for Potatoes [Online]. Available: https://www.ams.usda.gov/sites/default/files/media/Potato_Visual_Aids_5B1%5D.pdf.

[11] https://github.com/maxjfeldman/Feldman_hardware_control

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[13] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *Int. Conf. Learning Represent.*, San Diego, CA, USA, 7–9 May 2015

[14] R.E. Nylund, J.M. Lutz, "Separation of hollow heart potato tubers by means of size grading, specific gravity, and X-ray examination," *American Journal of Potato Research* 1950, 27, pp. 214-222, doi: 10.1007/BF02850188

[15] E.E. Finney Jr., K.K. Norris, "X-ray scans for detecting hollow heart in potatoes," *American Potato Journal* 1978, 55, pp. 95-105, doi: 10.1007/BF02852096

[16] D.C. Nelson, M.C. Thoreson, D.A. Jones, "The relationship between tuber size and hollow heart in Norgold Russet potatoes," *American Potato Journal* 1979, 56, pp. 329–337, doi: 10.1007/BF02853848

[17] G.S. Birth, "A nondestructive technique for detecting internal discolorations in potatoes," *American Potato Journal* 1960, 37, pp. 53-60, doi: 10.1007/BF02855060

[18] R.B. Harvey, "The X-ray inspection of internal defects of fruits and vegetables," *Journal of the American Society for Horticultural Science* 1937, 35, pp. 156–157

[19] E.E. Finney Jr., K.K. Norris, "X-ray images of hollow heart potatoes in water," *American Potato Journal* 1973, 50, pp. 1–9, doi: 10.1007/BF02851513

[20] A. Dacal-Nieto, A. Formella, P. Carrión, E. Vazquez-Fernandez, M Fernández-Delgado, "Non–destructive Detection of Hollow Heart in Potatoes Using Hyperspectral Imaging," *In: P. Real, D. Diaz-Pernil, H. Molina-Abril, A. Berciano, W. Kropatsch. (eds) Computer Analysis of Images and Patterns. CAIP 2011. Lecture Notes in Computer Science,* vol 6855, Springer, Berlin, Heidelberg, doi: 10.1007/978-3-642-23678-5_20

[21] K.C. Watts, L.T. Russell "A review of techniques for detecting hollow heart in potatoes," *Canadian Agricultural Engineering* 1985, 27, pp. 85-90