Development of a Deep-Learning Pipeline to Detect and Locate Contaminants of Industrial Products via non-Invasive Microwave Signals

Margherita Musumeci[‡], Juan Sebastian Amaya Cano[§], Filippo Lazzati[‡], Chiara Martano[¶], Francesco Pappone [‖], Claudio Ramonda [¶], Marco Ricci[**], Jorge A. Tobon V.[*], Giovanna Turvani[*], Mario R. Casu[*], Marco Mussetta [†], and Francesca Vipiana[*]
[*]Dept. of Electronics and Telecommunications, Politecnico di Torino, Turin, Italy
[¶]Dept. of Management and Production Engineering, Politecnico di Torino, Turin, Italy
[‖]Dept. of Mathematical Sciences, Politecnico di Torino, Turin, Italy
[†]Dept. of Energy, Politecnico di Milano, Milan, Italy
[‡]Dept. of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy
[§]Dept. of Chemistry, Materials and Chemical Engineering, Politecnico di Milano, Milan, Italy,
[**] Wavision Srl, Antares Vision group, Turin, Italy
emails: francesca.vipiana@polito.it, marco.mussetta@polimi.it

***Abstract*—Nowadays, food companies produce large volumes of packaged food products to satisfy the demand of a population that keeps increasing in number. One of the biggest challenges these enterprises must face is how to scale accurate contaminant detection methods. In this paper, we improve atop existing microwave-sensing techniques for food contaminant detection by adding an ensable of Machine Learning (ML) techniques. We consider a greater variety of common contaminants in the food industry with respect to the literature. Moreover, we enhance the data collection phase and propose a Graph Neural Network (GNN)-base approach to detect the position of the contaminant. We show that this enhanced data combined with the proposed ensemble of ML algorithms outperforms the accuracy of the detection with respect to the state-of-the-art approaches.**

***Index Terms*—microwave sensing, machine learning, imaging, graph neural networks, food technology**

I. Introduction

A food company's reputation relies on customer trust, and any foreign bodies in packaged products can harm this trust and brand loyalty. Ensuring consumer health and preventing food contamination is crucial in multi-stage production processes like the food industry [1]. Food contaminant detection systems, such as X-Ray imagers, metal detectors, etc., are used to prevent such issues, but they may have limitations, especially for contaminants with low density. In Tab I, a list of the most commonly used detection technique is presented, together with their characteristics and limitations. To address this problem, a new approach called ML-based Microwave Sensing (ML-MWS) is proposed in [2]. Thanks to the use of Microwave-based sensing, it shows that it is possible to use Machine Learning algorithms to achieve a high detection accuracy on some contaminants and to run detection in real-time. Next, the entire procedure for the development and deployment of such kind of system has been rigorously defined in [3]. Another related work [4] introduces a custom neural pattern recognition network, achieving near-optimal accuracy with limited features and samples.

Building upon the previous research, this paper aims to improve detection results by enhancing data collection, using an ensemble of ML algorithms, and leveraging imaging techniques to detect the contaminant's position. Additionally, the research expands the variety of products considered, leading to improved detection accuracy across a wider range of contaminants with fewer samples.

TABLE I
USE CASES AND LIMITATIONS OF DETECTION TECHNIQUES USED IN MANUFACTURING.

Detection system	**Low-density contaminants detection**	**Need for trained operators**	**Safety issue for operators**	**Ability to see through metal**
Microwaves	Yes	No	No	No
X-Rays	No	Yes	Yes	Yes
Near-Infrared	No	No	No	No
Metal Detectors	No	No	No	No
Visual Inspection	No	No	No	No

II. Related Works

As per the classification provided by the FDA, the hazards can be categorised based on their inherent properties; including

This work was supported in part by the project PON Research and Innovation "Microwave Imaging and Detection powered by Artificial Intelligence for Medical and Industrial Applications (DM 1062/21)", funded by MUR. It was carried out partially within the Agritech National Research Center, funded by the European Union Next-GenerationEU (Piano Nazionale di Ripresa e Resilienza (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4 – D.D. 1032 17/06/2022, CN00000022), and supported by Alta Scuola Politecnica through Wavision project.

physical attributes such as plastic, wood, and glass; chemical components such as heavy metals; and biological factors like listeria and salmonella [5]. Physical hazards represent a threat to consumer well-being and manufacturing equipment integrity, thus improving their detection constitutes a crucial element of a comprehensive food safety program. At the current state, mass spectrometry [6] and Raman spectroscopy [7] are shown to be effectively combined with ML to assess the quality of food against adulteration (in white rice in the cited paper) and to detect food-born pathogens, respectively. Microwaves have been used to sense the moisture content in corn in combination with deep neural networks [8], then, following the successful application of MWI to the medical imaging field [9], there has been an increasing interest toward industrial applications of this technique [10]. The authors of [11] developed a hand-held time-domain reflectometer working in the microwave spectrum to assess food quality by measuring variations in the dielectric properties, which can be determined by a variation of water concentration. As a result of an industrial implementation [12], a radar-based detection system reveals the ability to detect foreign bodies such as wood, plastic, bone, and fruit stones but is fundamentally different from our approach as it applies to pipes where liquid food or emulsions can flow before being packaged. Conversely, in [13] MW sensing is used to detect plastic contaminants in packaged cheese slices. In this case, a single antenna patch is used to illuminate the target and the reflected signal is shown to be sufficiently modified by the contaminant to allow for easy detection. Such a manageable condition does not apply in the case study shown in [14], in which the acquired signals in the contaminated and uncontaminated cases do not exhibit immediately visible patterns useful for discrimination, hence requiring further investigation of more efficient techniques for more automated pattern recognition.

III. Benchmark Dataset

In order to improve the ability of the algorithm to detect contaminants, we decided to broaden the existing dataset on which the algorithm is based, which was first built in [3]. The goal was to have a model that is able to adapt to different products with very different behavior when it comes to their "dielectric properties". Therefore, we selected a carbonated soft drink (soda), a soft drink with low amounts of CO_2 (ice tea), soy sauce, flour and honey. The chosen contaminants were plastic, paper, wood, glass, aluminum, glue, and cork, commonly found in packaging according to [1].

To build the machine's detection efficiency, we initially scanned uncontaminated products, some of which were opened and closed to mimic the pressure conditions of contaminated ones. Then, we proceeded to select the contaminant positions based on the behavior of the products. For liquids like soft drinks, contaminants would either float or sink. For higher-density products like flour, eight different surface, middle, and bottom positions were considered for measurements (see Tab. II).

TABLE II
Products and contaminants sizes included in our dataset. (P) stands for paper container while (G) is for glass container.

Contaminant	Product					
	Soda	Flour (P)	Flour (G)	Honey	Soy	Tea
Clear	500	100	200	200	300	300
Plastic	160	-	240	80	160	120
Paper	-	160	240	-	-	-
Wood	40	-	-	-	40	-
Glass	100	-	-	80	120	-
Aluminium	140	-	-	-	60	-
Glue	-	160	-	-	-	-
Cork	40	-	-	-	40	-
Tot	**980**	**420**	**680**	**360**	**680**	**420**

Fig. 1 shows the entire MWS system employed. The MWS system consists of six PCB-printed monopole antennas arranged in an arch above the production line, allowing uninterrupted product flow. Each of the six antennas involved in the setup is directly connected to a 6-port VNA through individual ports. Experimental data is gathered by interfacing a laptop with the VNA, serving as the controller during the data collection process. To minimize interference and reflections, a shielding case encloses the antenna arch. The Control Center is responsible for the activation and the speed of the line.

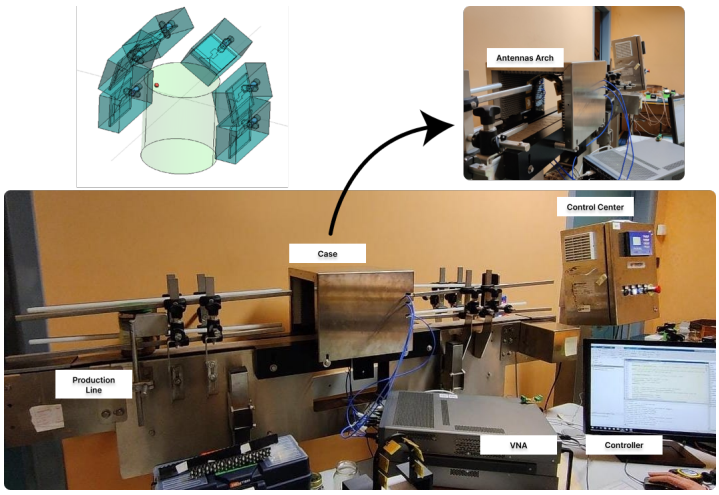

Fig. 1. Microwave Sensing system utilized in our experiments.

IV. Machine Learning Classification

Data collection is the first step in the process of ML-based Microwave Sensing. Next, we apply a ML algorithm for binary classification, i.e., to detect whether the food product is contaminated or not. The features used to feed to the ML algorithm are the same as defined in [2]: a 6×6 complex scattering matrix for 11 evenly spaced frequencies. The measurements are performed in a bandwidth of frequencies that ranges from 4 to 6 GHz for the flour and from 3 to 5 GHz for the other products. Since each of these elements is a complex number with real and imaginary parts, each sample in the

Dataset is a vector of $matrix \times matrix \times frequencies \times real/imaginary = 6 \times 6 \times 11 \times 2 = 792$ features.

A. Issues of previous implementations

The neural network implementation provided in [4] suffers from some shortcomings. In particular, it does not achieve the same generalization accuracy on different types of products and contaminants. Indeed, as anticipated, we extended the dataset to a larger variety of products and contaminants, and testing the same network architecture of the work in [15] with these products, the ML algorithm fails in achieving 100% accuracy (see Tab. III). As the main goal for food companies is to avoid contaminated products go to market, it is important to minimize the number of *False Negatives*, thus an alternative ML implementation is here presented.

TABLE III
TEST ACCURACY OF [2] ARCHITECTURE. IN BLUE IS REPORTED THE RESULT MENTIONED [15].

Product	Accuracy	FN
Hazelnut-Cocoa Cream	100%	0
Soda	93.9%	10
Flour	100%	0
Honey	100%	0
Soy	95.8%	8
Tea	89.3%	12

B. ML Algorithms comparison

A key goal of this implementation consists in the optimization of the time required for both the training procedure and the classification task, respectively to minimize the "calibration" time for a novel implementation, and further to support a real-time processing. A variety of ML algorithms in addition to Neural Networks were analyzed, to understand which solution can give the required time performance while also keeping the highest possible degree of accuracy (See Fig. 2).

Lasso regression implements linear regression with ℓ_1-norm regularization. It outperforms other methods, showing the importance of regularization. It is a simple model that is fast to train and can provide real-time predictions. Another promising algorithms are Ada Boost [16], which exploits boosting techniques to improve the accuracy of the dataset. Good methods are also the ensembles like Bagging Trees and Random Forests. Fig. 2 shows that also a lazy classifier like k-Nearest Neighbors can reach high accuracy, but suffers from slow prediction time. To improve the detection accuracy, we decide to adopt an ensemble of the best methods, since, under the assumption of independence, we can model the prediction process as a Binomial distribution, and considering majority voting we have that error can be reduced thanks to the formula:

$$\mathbb{P}\big(X \geq \lceil \frac{n}{2} \rceil \,\big|\, p = \epsilon\big) = \sum_{i=\lceil \frac{n}{2} \rceil}^{n} \binom{n}{i} \epsilon^i (1-\epsilon)^{n-i} < \epsilon \quad \text{if } \epsilon < 0.5$$

where X is the random variable saying the number of predictors with error probability ϵ over the total number n of predictors that give a wrong prediction. Clearly, if $\epsilon > 0.5$ (as seen in Fig. 2), then the error probability of the ensemble method reduces. So, we choose to use an ensemble of Lasso and Ada Boost regression, which shows the best performance (see Tab. IV).

Consider Tab. IV: in terms of accuracy, it is clear that Lasso classifier is outperforming all other models. It achieves almost-optimal accuracy in all the considered products. However, let us strengthen the idea that we prefer to identify as contaminated some additional good products, if we are able to reduce the number of contaminated products classified as intact, in other words, reducing the False Negatives cases. To do so, let's take a look at Tab. V. Most algorithms correctly classify *honey* and *flour*. By considering other products, we have selected the ensemble of Lasso and AdaBoost classifiers as the best one. The choice of these specific algorithms is based on both performance and computational reasons. Other algorithms require too much time to automatically tune the hyperparameters or provide worse performances. This choice balances this trade-off.

C. Hyperparameters Tuning and Model Selection

To tune the hyperparameters, we adopted Bayesian Optimization and Grid search methods which helped us to find the best hyperparameters. Consequently, this approach led to the development of slightly varied models for different products included in our study. We adopted nested k-fold cross-validation to choose the parameters. We use a standard train-test split of the dataset to train and validate the specific model (80% train, 20% test).

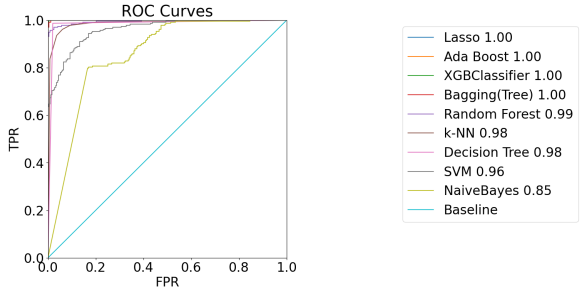

Fig. 2. Comparison of ROC curves for different ML algorithms on Hazelnut-Cocoa Cream dataset.

V. CONTAMINANT LOCALIZATION WITH GRAPH NEURAL NETWORKS

The main scope of inspection devices is to identify contaminated samples. But also, knowing extra information about the contaminant could result in precious information. For example, the composition and the position of the intrusion could help in identifying the sources and the reasons for the contamination.As it can be easily deduced by the matrix structure of our input signal, our problem can be naturally interpreted as a graph problem ([17], [18]): the scattering matrix $\mathbf{S}$ can be seen as the adjacency matrix of a graph whose nodes v_i are represented by electrodes. Graph Neural

TABLE IV
EXPERIMENT RESULTS REPORTED WITH ACCURACY METRIC. OUR *Baseline* ARE THE RESULTS REPORTED IN TAB. III. ENSEMBLE IT'S A COMBINATION OF LASSO AND ADABOOST.

Model	**Product**				
	Soda	Honey	Flour (P)	Soy	Tea
Baseline	93.9%	100%	100%	95.9%	89.3%
Lasso	**100%**	**100%**	**100%**	**97.6%**	**99.0%**
AdaBoost	**99.0%**	99.4%	**100%**	94.8%	93.3%
Bagging Tree	93.7%	**100%**	**100%**	95.8%	84.8%
Decision Tree	**94.2%**	99.7%	**100%**	94.2%	81.0%
Random Forest	92.4%	**100%**	**100%**	94.2%	82.4%
Naive Bayes	84.4%	**100%**	99.5%	88.9%	66.7%
1-hidd. layer NN	**97.3%**	**100%**	98.8%	**96.6%**	**94.3%**
3-hidd. layers NN	**96.5%**	**100%**	**100%**	94.5%	89.0%
Ensemble	**99.4%**	99.4%	**100%**	93.5%	**96.6%**

TABLE V
EXPERIMENT RESULTS OF THE MOST PROMISING METHODS REPORTED WITH THE NUMBER OF FALSE NEGATIVE OUR *Baseline* ARE THE RESULTS REPORTED IN TAB. III. ENSEMBLE IT'S A COMBINATION OF LASSO AND ADABOOST.

Model	**Product**				
	Soda	Honey	Flour (P)	Soy	Tea
Baseline	10	0	0	8	12
Lasso	1	**0**	**0**	7	2
AdaBoost	4	1	**0**	11	14
Mini-NN	8	**0**	1	20	10
Deep-NN	24	**0**	1	22	20
Ensemble	**0**	**0**	**0**	**6**	**0**

Networks rely, in general, on a *message passing* architecture ([18]), which can be summarized as a weighted diffusion of a combination of

- *Node embeddings*, which we will denote as $\mathbf{z_i}^{(k)} \in \mathbb{R}^n$, referring to the $k-th$ embedding, resulting from the $k-th$ message passing iteration, of the $i-th$ node
- *Edge embeddings*, which we will denote as $\mathbf{e_{ij}}^{(k)} \in \mathbb{R}^n$, referring to the $k-th$ embedding, resulting from the $k-th$ message passing iteration, of the e_{ij} edge
- *Graph embeddings*, which we will denote as $\mathbf{z}_{\mathcal{G}}^{(k)} \in \mathbb{R}^n$, referring to the $k-th$ embedding, resulting from the $k-th$ message passing iteration, entire graph $\mathcal{G}$

A. Implementation

Our implementation relies on the *Pytorch Geometric* framework. Let $\mathbf{S}^{(t)} \in \mathcal{M}_{6\times6}(\mathbb{C})$ and $s_{ij}^{(t)}$ its components, for $t \in 0, 1 \cdots, 10$ representing the 11 acquisitions executed per measure. Let $\mathbf{s}_{ij} \in \mathbb{C}^{11}$ indicate the vectors containing the 11 measurements of a signal from electrode i by electrode j. Initial embeddings are produced as follows:

- $\mathbf{z_i}^{(0)} = \mathbf{z}_i^{pos} \oplus \mathcal{R}e(\mathbf{s}_{ii}) \oplus \mathcal{I}m(\mathbf{s}_{ii})$
- $\mathbf{e}_{ij}^{(0)} = l_{ij} \oplus \mathcal{R}e(\mathbf{s}_{ij}) \oplus \mathcal{I}m(\mathbf{s}_{ij})$

with l_{ij} being the euclidean distance between electrodes i and j, and $\mathbf{z_i^{pos}}$ being the normalized, 2D polar coordinates for node i. The message passing architecture is GENconv, from [19]. A single GENconv layer is utilized to produce the embeddings. Our architecture performs a double classification task, producing two distinct output vectors ($\mathbf{y}_{cont}$ and $\mathbf{y}_{pos}$) each normalized with a softmax layer. A custom loss summing Crossentropy losses for both classification tasks (contaminant and location) is employed. The datasets are specifically labeled to encode information about the approximate position of the contaminant in the container. Depending on the dataset, different positions are used. The location component is hence effectively a classification over the possible positions.

B. Results

Models are trained over nested 5-fold Cross Validation on the training set, for a total of 200 epochs of training for each step of the Bayesian optimization pipeline (with a total of 50 optimization steps). The final training (on both the previous validation and training splits) is then 1000 epochs long. The total data split is hence: Training composed by 64% of the dataset, while validation is 16% and testing 20%. results can be seen in Table VI.

TABLE VI
DATASET STATISTICS FOR THE TRAINED GENCONV MODELS. *N.positions* REFERS TO THE NUMBER OF POSSIBLE POSITIONS AVAILABLE IN THE DATASET. *N. contaminants* REFERS TO THE NUMBER OF DIFFERENT CONTAMINANTS. MODELS TRAINED ONLY ON CONTAMINANT CLASSIFICATION HAVE N.A. IN THEIR ASSOCIATED *N. positions*.

Dataset	**N. Data**	**N. Positions**	**N.Contaminants**	**Acc.**
Soy	620	N.A.	5	95.43%
Tea	420	N.A.	1	81.48%
Flour (G)	680	18	2	98.29%
Flour (P)	420	12	2	89.51%
Soda	980	N.A.	5	89.33%
Honey	360	2	3	94.36%

VI. LIMITATION AND CHALLENGES

The work has some limitations to take into account. Although the system performs well in the laboratory's MWS system, its real-world implementation require addressing challenges such as limited computational resources and diverse environmental conditions. Indeed, the ML ensemble developed may not be easily deployed on certain hardware platforms. The study's scope of contaminants and food products might not fully represent what food companies encounter, necessitating additional research and validation as well as different packaging materials.

VII. CONCLUSION AND PERSPECTIVES

In this study, training and testing alternative ML algorithms on the new dataset that we collected, we demonstrated that a Lasso Regressor ensemble with Ada Boost can often outperform the neural networks in contaminants detection. Additionally, we introduced the utility of GNN for broader classification tasks and contaminant position reconstruction. Future research should consider extending the dataset with additional food products and contaminants, to further reduce the dataset size used by ML algorithms for detection and to develop a ML algorithm capable of generalizing to unseen products and contaminants.

REFERENCES

[1] J. Litster and I. Bogle, "Smart process manufacturing for formulated products," *Engineering*, vol. 5, no. 6, pp. 1003–1009, 2019.

[2] L. Urbinati, M. Ricci, G. Turvani, J. A. T. Vasquez, F. Vipiana, and M. R. Casu, "A machine-learning based microwave sensing approach to food contaminant detection," in *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1–5, 2020.

[3] M. Ricci, B. Štitić, L. Urbinati, G. D. Guglielmo, J. A. T. Vasquez, L. P. Carloni, F. Vipiana, and M. R. Casu, "Machine-learning-based microwave sensing: A case study for the food industry," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 11, no. 3, pp. 503–514, 2021.

[4] A. Darwish, M. Ricci, F. Zidane, J. A. T. Vasquez, M. R. Casu, J. Lanteri, C. Migliaccio, and F. Vipiana, "Physical contamination detection in food industry using microwave and machine learning," *Electronics*, vol. 11, no. 19, 2022.

[5] "Hazard analysis and risk-based preventive controls for human food: Draft guidance for industry," 2022.

[6] D. K. Lim, N. P. Long, C. Mo, Z. Dong, L. Cui, G. Kim, and S. W. Kwon, "Combination of mass spectrometry-based targeted lipidomics and supervised machine learning algorithms in detecting adulterated admixtures of white rice," *Food Research International*, vol. 100, pp. 814–821, 2017.

[7] S. Yan, S. Wang, J. Qiu, M. Li, D. Li, D. Xu, D. Li, and Q. Liu, "Raman spectroscopy combined with machine learning for rapid detection of food-borne pathogens at the single-cell level," *Talanta*, vol. 226, p. 122195, 2021.

[8] J. Zhang, D. Du, Y. Bao, J. Wang, and Z. Wei, "Development of multifrequency-swept microwave sensing system for moisture measurement of sweet corn with deep neural network," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 9, pp. 6446–6454, 2020.

[9] R. Chandra, H. Zhou, I. Balasingham, and R. M. Narayanan, "On the opportunities and challenges in microwave medical sensing and imaging," *IEEE Transactions on Biomedical Engineering*, vol. 62, pp. 1667–1682, July 2015.

[10] Z. Wu and H. Wang, "Microwave tomography for industrial process imaging: Example applications and experimental results," *IEEE Antennas and Propagation Magazine*, vol. 59, no. 5, pp. 61–71, 2017.

[11] O. Schimmer, F. Daschner, and R. Knochel, "Uwb-sensors in food quality management — the way from the concept to market," in *2008 IEEE International Conference on Ultra-Wideband*, vol. 2, pp. 141–144, 2008.

[12] M. Edwards, *Detecting foreign bodies in food*. Woodhead Publishing Ltd, 2004.

[13] M. Muradov, P. Kot, M. Ateeq, B. Abdullah, A. Shaw, K. Hashim, and A. Al-Shamma'a, "Real-time detection of plastic shards in cheese using microwave-sensing technique," *Proceedings*, vol. 42, no. 1, 2020.

[14] H. Bisgin, T. Bera, H. Ding, H. G. Semey, L. Wu, Z. Liu, A. E. Barnes, D. A. Langley, M. Pava-Ripoll, H. J. Vyas, *et al.*, "Comparing svm and ann based machine learning methods for species identification of food contaminating beetles," *Scientific reports*, vol. 8, no. 1, pp. 1–12, 2018.

[15] M. Ricci, J. A. T. Vasquez, G. Turvani, I. Sirena, M. R. Casu, and F. Vipiana, "Microwave sensing for food safety: a neural network implementation," in *2021 IEEE Conference on Antenna Measurements Applications (CAMA)*, pp. 444–447, 2021.

[16] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006.

[17] M. M. Bronstein, J. Bruna, T. Cohen, and P. Veličković, "Geometric deep learning: Grids, groups, graphs, geodesics, and gauges," *arXiv preprint arXiv:2104.13478*, 2021.

[18] W. L. Hamilton, R. Ying, and J. Leskovec, "Representation learning on graphs: Methods and applications," *arXiv preprint arXiv:1709.05584*, 2017.

[19] G. Li, C. Xiong, A. Thabet, and B. Ghanem, "Deepergcn: All you need to train deeper gcns," 2020.