Bio-FP: Biochip Fingerprints for Authentication

Navajit Singh Baban, Sohini Saha, Sofija Jancheska, Jiarui Zhou, Sanjairaj Vijayavenkataraman, Sukanta Bhattacharjee, Yong-Ak Song, Krishnendu Chakrabarty and Ramesh Karri

2023 IEEE Biomedical Circuits and Systems Conference (BioCAS) | 979-8-3503-0026-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/BioCAS58349.2023.10388991

***Abstract*—Microfluidic biochips are widely used in biological computing, clinical diagnostics, and point-of-care tests. However, the growing demand and the complex supply chain of biochips expose them to intellectual property (IP) attacks such as counterfeiting, overbuilding, and piracy. To address this issue, we present a biochip-level fingerprinting (Bio-FP) scheme. We utilize melt-electrospinning printing technique to print unique Bio-FPs directly onto biochips. Then, a layer of polydimethylsiloxane (PDMS) is applied through spin-coating to obscure the Bio-FPs. If the Bio-FPs are doped with a fluorescent dye, they can be detected by shining UV light. Authentication of dyed Bio-FPs is achieved through spectral analysis by mapping the intensity-wavelength response. To optimize the authentication scheme for Bio-FPs, several pre-processing techniques were employed to enhance their quality. Additionally, transfer learning and fine-tuning were utilized with multiple deep learning models, yielding a high Bio-FP classification accuracy of 95.8%.**

***Index Terms*- biochip, fingerprints, melt electrospinning, deep learning, intellectual property, counterfeiting.**

I. Introduction

Microfluidics is an interdisciplinary field that focuses on manipulating fluids at very small volumes, typically microliters or nanoliters [1]. A microfluidic biochip, also known as a lab-on-a-chip, combines various biochemical functionalities into a single miniaturized device, mimicking the capabilities of a laboratory [2]. Compared to traditional bench-top laboratories, biochips excel in dispensing, mixing, splitting, and transportation, thanks to the small sample sizes involved [2]. They have revolutionized biological computing in numerous areas, including enzymatic, DNA, and proteomic analysis, genetic and polymerase chain reaction (PCR) studies, surface immunoassays, and toxicity monitoring [2].

By 2025, the global microfluidics market is projected to grow from its estimated value of $15.7 billion in 2020 to a substantial size of $44 billion [3]. However, as with most emerging technologies, innovation takes precedence and security is a secondary consideration, only addressed after vulnerabilities are identified. For example, a staggering 40 million worth of counterfeit or substandard COVID test kits has been confiscated in 77 countries, with 407 individuals arrested between December 2019 and June 2020 [4].

Biochip companies have been adopting horizontal supply-chain models to achieve economies of scale and cost reduction [2], [5]. Untrusted third parties in the supply-chain pose risks of intellectual property (IP) attacks including reverse engineering, counterfeiting, and overbuilding [2], [5]. We propose a Bio-FP, a biochip fingerprint scheme to authenticate biochips. Each biochip is marked with a unique fingerprint using a melt-electrospinning printer. The fingerprints are concealed with a layer of polydimethylsiloxane (PDMS). By doping the ink with a fluorescent dye, the fingerprints can be identified when exposed to UV light. Spectral analysis of the fingerprints provides a unique intensity-wavelength response mapping. The scheme offers two layers of authentication: hidden fingerprint and unique spectral response. We used deep learning (DL) to distinguish authentic and counterfeit fingerprints. We use transfer learning with pre-trained DL models to ensure robust and accurate authentication. Training on our Bio-FP dataset yielded algorithms with high accuracy of 95.8%, offering a reliable solution for authentication.

N.S. Baban, J. Zhou, S. Vijayavenkataraman, and Y. Song are with the Department of Engineering, New York University Abu Dhabi, UAE.

S. Jancheska and R. Karri are with the Department of Electrical and Computer Engineering, New York University, NY, USA.

S. Bhattacharjee is with the Department of Computer Science and Engineering, Indian Institute of Technology Guwahati, India.

S. Saha is with the Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA.

K. Chakrabarty is with the School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ, USA.

II. Background

A. IP-based Threats on Biochips

The manufacturing process of biochips has several stages and various entities, some of which may be untrustworthy. This can result in IP theft through reverse engineering. An attacker can access valuable information such as the biochip's architecture, materials, functions, and bio-protocol by reverse engineering a biochip. Adversaries can use this information to engage in IP piracy, counterfeiting, and over production of biochips.

Fig. 1 illustrates the threat model. The cyber-physical biochip has a physical biochip, an end user, and a biochip company (or trusted third party, TTP). The biochip company or TTP authenticates the biochip using a DL-based classifier, with access to challenge-response pair database (CRPDB) for each Bio-FP. Equipped with a camera, the company or TTP can use the Bio-FP image for biochip authentication. The end user transmits the Bio-FP image to an isolated TTP through

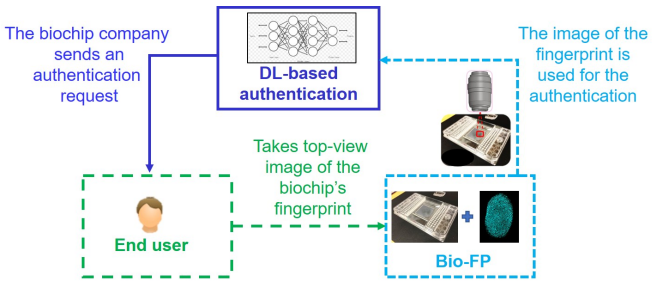

Fig. 1: IP-based threats on Biochips.

alternate network interfaces. The Bio-FP is the authentication key, enabling the company/TTP to verify provenance.

B. Prior Work on IP Protection of Biochips

Protecting biochips from IP-based attacks has focused on watermarking [2], [6] and obfuscation [7]. Baban et al. proposed a watermarking scheme for FMBs, by increasing the height of micro reaction chambers or microchannels at specific positions to create fluorescent watermarks that can be measured using fluorescence microscopy A previous study uses molecular bar codes at the protocol level to safeguard the biochemical sample IP [8]. The scheme hierarchically embeds secret signatures, using mixing ratio, incubation time, and sensor calibration to protect the bio-samples. Thus far, no work has been done to secure biochips against IP-based threats via fingerprint authentication. This work provides a biochip-level scheme that produces unclonable fingerprints.

III. Biochip Fingerprinting (Bio-FP) Solution

We embed unique fingerprints on biochips (Bio-FPs). Fig. 2(a) shows the melt-electrospinning setup (RegenHu 3D Discovery bioprinter) to create a fingerprint. Polycaprolactone (PCL) pellets with a molecular weight of 45,000 are used. Print parameters include 85°C temperature, a distance of 6 mm between needle and substrate, 0.125 MPa pressure, and 8.0 kV voltage. These parameters form random spiral fibers in the fingerprints. The print head is moved to a specified position and discharged for 5 seconds [9].

A. Bio-FP process

Fig. 2(b) shows six fingerprints that were printed on PDMS circular samples using the above mentioned process parameters. Each fingerprint has subtle variations in structural design. Fig. 2(c) shows "NYU" printed using a bioprinter, concealed by spin coating a layer of PDMS (Fig. 2(d)). By integrating a fluorescent perylene dye into the ink (PCL) of the melt-electrospinning bioprinter, the fingerprints were made fluorescent, rendering them visible solely under UV light. Fig. 2(e) shows a schematic with ten 1 mm circles on a 30 mm diameter circular PDMS layer. When printed using the melt-electrospinning bioprinter, the print showed a random pattern (Fig. 2(f)). The mismatch between the ten circles and the printed fingerprint reveals the randomness associated with the process. It is unlikely for an attacker to reverse engineer the input print commands from the printed fingerprint.

To obfuscate the fingerprint, a layer of PDMS is coated (Fig. 2(g)). Finally, the sample was exposed to a 365 nm UV light to read the fingerprint (Fig. 2(h)). We implemented deep learning (DL) classifiers to distinguish between authentic from fake fingerprints. Our DL models were trained using 437 authentic and 253 fake fingerprints, which were printed on glass slides at equal intervals. The authentic fingerprints were printed in a single session with specified process parameters. The fake fingerprints were created by either using the same process parameters but at a different time or by using a different voltage setting other than the specified 8 kV.

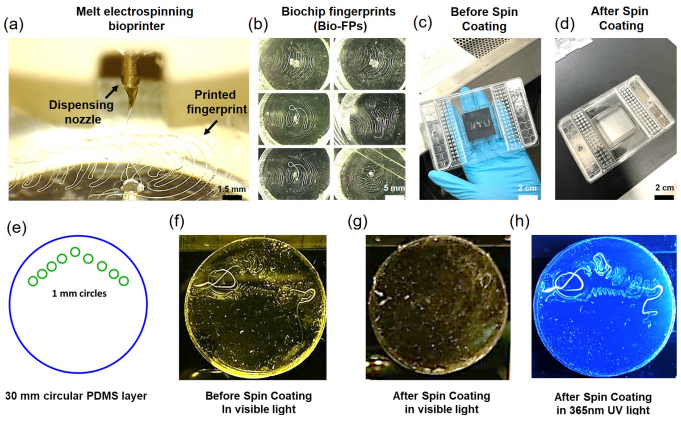

Fig. 2: (a) Melt-electrospinning bioprinter shows Bio-FP. (b) Six Bio-FPs use the same parameters. (c) "NYU" printed before spin coating with PDMS. (d) After spin coating, "NYU" is invisible. (e) Schematic shows ten 1 mm circles on a 30 mm diameter PDMS layer, (f) before spincoating with PDMS layer. (g) Printed circles are invisible after spin coating. (h) Printed circles visible with 365 nm wavelength UV light.

B. Bio-FP Analysis

To study if Bio-FPs are random, we examined whether their distribution adheres to a normal distribution. We obtained unique single-point values from Bio-FP images using two techniques. First, we obtained a single-point value representation from an image using the VGG16 (*vgg*) [10] ML model by applying global average pooling to the output feature maps. These averaged values are concatenated to create a compact feature vector that captures the essential image characteristics. Second, we converted the images to grayscale (*gray*) yielding a single-channel representation and computed the average intensity by summing the intensities of all pixels and dividing the sum by number of pixels.

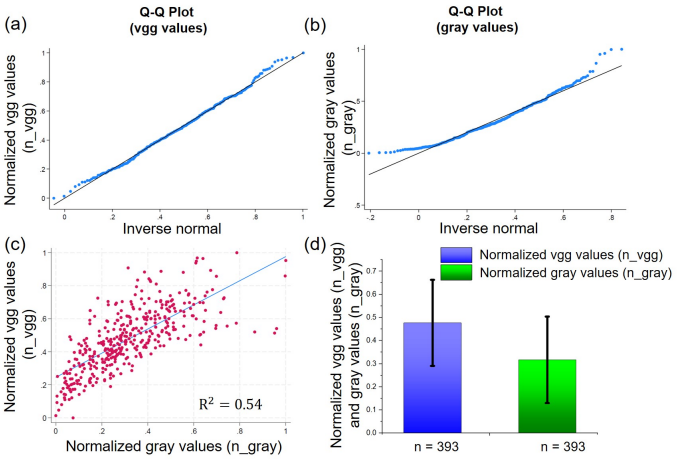

Fig. 3: (a) Quantile-quantile (Q-Q) plot for *n_vgg*. (b) Quantile-quantile (Q-Q) plot for *n_gray* values. (c) Two scatter plot and regression analysis for *n_vgg* and *n_gray*. (d) Mean and standard deviation bar graphs for *n_vgg* and *n_gray* values.

We normalized the values using minimum and maximum

values among each class to bring them between 0 and 1. The normalized vgg (*n_vgg*) and grayscale (*n_gray*) values were used in quantile-quantile (Q-Q) plots to check whether the values follow a normal distribution. Plots in Fig. 3(a) confirm that *n_vgg* values ($n = 393$) follow a normal distribution as the points fall along the straight line.

However, *n_gray* values ($n = 393$) do not follow a normal distribution as seen in Fig. 3(b). The criterion was also tested using skewness-kurtosis in STATA statistical package, yielding consistent results. We attribute this to the complexity of the data and the diverse characteristics of the computations. Grayscale values, representing an image processing technique, did not conform to a normal distribution despite the Bio-FPs exhibiting normal distribution. This emphasizes the efficacy of DL image analysis-based authentication.

We used a scatter plot and linear fit to estimate the correlation, showing a 54% (R-square value) relationship (Fig. 3(c)) between *n_vgg* and *n_gray* values. To measure relative variability of the dataset, we calculated coefficient of variation (CV) for Bio-FPs. CVs for *n_vgg* and *n_gray* were 40% (mean = 0.477, standard deviation = 0.186) and 59% (mean = 0.317, standard deviation = 0.187), respectively (Fig. 3(d)). High CVs in Bio-FPs confirm their stochastic nature.

IV. Deep Learning-based Defense using Bio-FP

In this section, we present a DL-based approach for Bio-FP classification. We begin with the data acquisition and preprocessing techniques employed. Then we delve into the classification methods used. The discussion includes the implementation of transfer learning and experimental settings employed to achieve optimal results. Finally, we showcase results of the models and compare the DL models.

A. Data Preprocessing

This is an essential step due to the inadequate quality of the original data, which is affected by noise. The noise originating from various sources such as image acquisition or transmission artifacts, distorts and creates inconsistencies in the images. Thus it is difficult to extract meaningful features and patterns for accurate classification. So we employ noise reduction. To enhance quality of the images and ensure reliable analysis, we also use binarization and thinning. Fig. 4(a) illustrates the approach and includes data acquisition and preprocessing steps (noise reduction, binarization, and thinning).

1) Noise Reduction: We remove the background of all images enabling us to eliminate the noise in the background due to differences in the microscope settings for the real and counterfeit images, such as the direction of the light and distance, affecting the thickness of the fingerprint curves and backgrounds. After denoising, we are left with the fingerprints.

2) Binarization: We transformed our dataset into binary images with red and blue colors. This enabled us to augment the initial dataset and simplify the data representation, allowing analysis of the pattern and features in the images.

3) Thinning: We used thinning to increase our dataset and enhance the features of the fingerprint. Thinning enhanced the clarity of images by transforming the binary regions into lines that look like the skeletons of those regions.

B. DL Classifiers

We consider five DL-based classifiers and assessed their performance on the Bio-FP dataset. The evaluation process helps to understand the advantages and constraints of each model. *1. DenseNet121* is a convolutional network where each layer is connected to every preceding layer, essentially receiving "collective knowledge". It shares its feature maps with all subsequent layers via concatenation. *2. MobileNetV2* uses depth-wise separable convolutions. This process comprises a depth-wise convolution followed sequentially by a point-wise convolution. *3. ResNet50* has 50 layers and uses residual learning, with shortcuts that facilitate information exchange. This mitigates performance degradation in deep networks and allows the network to learn complex features and deeper representations, enhancing image classification accuracy. *4. EfficientNetV2B0* balances accuracy with computational efficiency. It employs compound scaling, efficient block design, and stochastic depth. *5. NASNetMobile* is designed for image classification on mobile devices with limited computational resources. It uses neural architecture search to identify network architectures for specific tasks and uses depth-wise separable convolutions, repeated cells, and efficient model scaling.

C. Transfer Learning Improves DL-based Classifiers

Transfer Learning [11] allows us to capitalize on a related classification task to improve performance of our DL models. Using pre-trained models trained on the ImageNet [12], we can reap three benefits. First, transfer learning reduces the time and data needed for training, as we can start from a point where the models have learned valuable features. Second, it generalizes our models to new, unseen data, enabling them to perform well even in scenarios with limited labeled data. Finally, pre-trained models can tap into complex representations that they learn from the ImageNet dataset. Using ImageNet for pre-training our classifiers offers a basis for transfer learning. We capitalize on the rich/diverse knowledge learned by the models.

As presented in Fig.4(b) we evaluated effectiveness of transfer learning using pre-trained DL models trained on ImageNet. All pre-trained layers were adjustable, with the output of the final layer extracted and saved. For classification, we enhanced the model for dropout regularization (dropout rate of 0.8) and incorporated softmax layers as activation functions. Model parameters were optimized to minimize classification error using categorical cross-entropy loss as the objective during training. We used the Adam optimizer to update the model's weights using a learning rate of 0.0001. Bio-FP dataset with labeled fingerprint images was employed for training and validation. To assess performance and generalizability of the model, the dataset underwent an 80-20 split, where 80% of the data was allocated for training and the remaining 20% for validation. During training, a batch size of 50 was employed for each iteration. Early stopping was incorporated to detect the optimal stage during training where performance on the validation set exhibits deterioration or plateaus, indicating overfitting. Our criteria for early stopping were based on a training accuracy reaching 98%, ensuring that the model does not specialize on the training data.

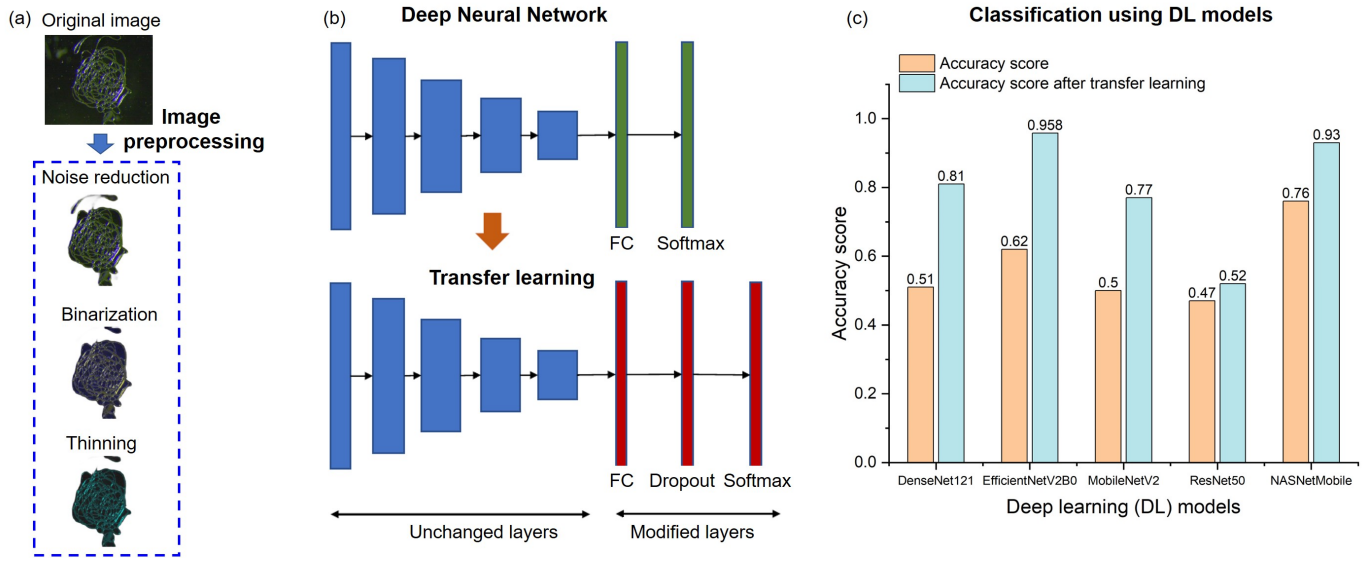

Fig. 4: DL Experiments: (a) DL pipeline includes image acquisition and pre-processing. (b) DL classifier using transfer learning to classify authentic and fake Bio-FPs. (c) Results of DL without and with transfer learning.

D. Experimental Results on DL-based Bio-FP Classification

We evaluated the trained models using a distinct validation dataset. We analyzed the results of transfer learning with pre-trained models alongside baseline DL models. The results are presented in Fig. 4(c), showcasing the performance of the DL models on the validation dataset. Transfer learning improves accuracy. EfficientNetV2B0 and NASNetMobile models had the highest accuracy scores of 95.8% and 93% respectively. Besides highlighting the effectiveness of transfer learning in improving performance, the results showcase the potential for more accuracy through pre-trained models.

DenseNet121, MobileNetV2, and ResNet50 did not yield good results compared to the models with transfer learning. DenseNet121 has 81% accuracy. MobileNetV2 and ResNet50 had lower accuracy. There could be many reasons behind this decrease. It is likely pre-trained weights of the models were ill-suited for our task or the target dataset had characteristics that differed from ImageNet on which the models were originally trained. Hence, the models struggled to adapt and learn meaningful representations for our task. The compound scaling ability of EfficientNetV2B0 allowed it to capture intricate patterns and features present in the data, resulting in its exceptional performance. Similarly, NASNetMobile, designed with an architecture derived from reinforcement learning, demonstrates its capacity to explore an extensive search space yielding optimal outcomes. In contrast, while DenseNet121, MobileNetV2, and ResNet50 are widely recognized and widely employed models, they exhibit limitations in terms of depth, network connectivity, or architectural design, contributing to their relatively lower performance.

V. Discussion and Conclusion

Biochip-fingerprints (Bio-FP) support authentication that operates at the biochip level. We used a melt-electrospinning printer to print Bio-FPs and enhance security by applying an invisible layer of PDMS. Bio-FPs can be detected using UV light and undergo spectral analysis for authentication. Data preprocessing, deep learning, and transfer learning increase the authentication accuracy to 95.8%. Bio-FPs, therefore, offer effective protection against IP-based attacks.

Some salient features of the Bio-FP scheme include the use of physically unclonable fingerprints, spin coating-based obscuration, and deep learning-based authentication. These features collectively contribute to the scheme's effectiveness in preventing forgery and tampering in biochips.

The scheme employs a melt electrospinning printer to generate physically unclonable fingerprints, making it challenging for attackers to forge them. The fingerprints' inherent unclonable nature prevents replication, even with knowledge of the process parameters. Additionally, a combination of spin coating-based obscuration and spectral analysis-based authentication techniques further enhances security and reduces the likelihood of successful forgery attempts.

In terms of tamper resistance, the scheme embeds the fingerprints beneath the biochip's surface using spin coating with a layer of polydimethylsiloxane (PDMS). This not only obscures the fingerprints but also physically secures them, making them inaccessible to potential attackers.

Deep learning methods play a pivotal role in the Bio-FP scheme's authentication process. By training deep neural networks, the scheme extracts compact and discriminative representations from fingerprint images, enabling accurate differentiation between genuine and forged fingerprints. Deep learning analysis encompasses various features, including texture and ridge structures, ensuring the scheme's reliability and performance.

In summary, the Bio-FP scheme's salient features, such as physically unclonable fingerprints, spin coating-based obscuration, and deep learning-based authentication, provide robust protection against forgery and tampering in biochips.

REFERENCES

[1] X. Xie, T. Gjorgjieva, Z. Attieh, M. M. Dieng, M. Arnoux, M. Khair, Y. Moussa, F. Al Jallaf, N. Rahiman, C. A. Jackson *et al.*, "Microfluidic nano-scale qpcr enables ultra-sensitive and quantitative detection of sars-cov-2," *Processes*, vol. 8, no. 11, p. 1425, 2020.

[2] N. S. Baban, S. Saha, A. Orozaliev, J. Kim, S. Bhattacharjee, Y. Song, R. Karri, and K. Chakrabarty, "Structural attacks and defenses for flow-based microfluidic biochips," *IEEE Transactions on Biomedical Circuits and Systems.*, vol. 16, no. 6, pp. 1261–1275, 2022.

[3] Markets and Markets, "Microfluidics market by product (devices, components (chips, sensors, pump, valves, and needles), application (ivd [poc, clinical, veterinary], research, manufacturing, therapeutics), end user and region - global forecast to 2025," 2020. [Online]. Available: https://www.marketsandmarkets.com/Market-Reports/microfluidics-market-1305.html

[4] "Fake COVID-19 kits seized in international trafficking crackdown from 77 countries," july 2020. [Online]. Available: https://www.ndtv.com/world-news/coronavirus-fake-covid-19-kits-seized-in-international-trafficking-crackdown-from-77-countries-2267147

[5] S. S. Ali, M. Ibrahim, J. Rajendran, O. Sinanoglu, and K. Chakrabarty, "Supply-chain security of digital microfluidic biochips," *Computer*, vol. 49, no. 8, pp. 36–43, 2016.

[6] M. Shayan, S. Bhattacharjee, J. Tang, K. Chakrabarty, and R. Karri, "Bio-protocol watermarking on digital microfluidic biochips," *IEEE Trans. Inf. Forensics Secur.*, vol. 14, no. 11, pp. 2901–2915, 2019.

[7] M. Shayan, S. Bhattacharjee, A. Orozaliev, Y.-A. Song, K. Chakrabarty, and R. Karri, "Thwarting bio-ip theft through dummy-valve-based obfuscation," *IEEE Trans. Inf. Forensics Secur.*, vol. 16, pp. 2076–2089, 2020.

[8] T.-C. Liang, K. Chakrabarty, T. Abaffy, H. Matsunami, and R. Karri, "Securing biochemical samples using molecular barcoding on digital microfluidic biochips," in *Proc. BioCAS*, 2021, pp. 01–06.

[9] N. S. Baban, "Bio-FP video using a melt-electrospinning bioprinter," Online on Youtube, June 2023, available: https://youtu.be/-643gDvnN5M [Accessed on 2023-06-01].

[10] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[11] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.

[12] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," *Proc. CVPR*, pp. 248–255, 2009.