Transfer Contrastive Learning for Raman Spectroscopy Skin Cancer Tissue Classification

Zhiqiang Wang, Yanbin Lin, Andrew C. Terentis, John Strasswimmer, and Xingquan Zhu, *Fellow, IEEE*

***Abstract*—Using Raman spectroscopy (RS) signals for skin cancer tissue classification has recently drawn significant attention, because of its non-invasive optical technique, which uses molecular structures and conformations within biological tissue for diagnosis. In reality, RS signals are noisy and unstable for training machine learning models. The scarcity of tissue samples also makes it challenging to learn reliable deep-learning networks for clinical usages. In this paper, we advocate a Transfer Contrasting Learning Paradigm (TCLP) to address the scarcity and noisy characteristics of the RS for skin cancer tissue classification. To overcome the challenge of limited samples, TCLP leverages transfer learning to pre-train deep learning models using RS data from similar domains (but collected from different RS equipments for other tasks). To tackle the noisy nature of the RS signals, TCLP uses contrastive learning to augment RS signals to learn reliable feature representation to represent RS signals for final classification. Experiments and comparisons, including statistical tests, demonstrate that TCLP outperforms existing deep learning baselines for RS signal-based skin cancer tissue classification.**

***Index Terms*—Transfer learning, contrastive learning, Raman spectroscopy, skin cancer, tissue classification.**

I. Introduction

SKIN cancer, characterized as the most prevalent cancer in the United States, presents a significant public health challenge. The statistics show that one in five Americans will develop skin cancer during their lifetime while one in four is impacted by skin disease [1], [2]. Nonmelanoma skin cancers (NMSC), including basal cell carcinoma (BCC) and squamous cell carcinoma (SCC), are particularly widespread, affecting over three million Americans annually [3]. The surge in skin cancer cases is not only in the U.S. but reported globally [4]. Melanoma and Keratinocyte skin cancer (KSC), which is considered strongly related to exposure to ultraviolet (UV) radiation, are the most common types in white populations, while the incidence of KSC, comprising BCC and SCC, far exceeds that of melanoma [5].

Predominantly, surgical intervention is a conventional approach to managing skin cancer, while it poses numerous challenges and bottlenecks at the same time. Differentiating cancerous from normal skin tissue during surgery is often difficult, and the procedures can be time-consuming, risky, and financially burdensome for patients [6]. In this context, RS emerges as a promising alternative. Preliminary studies have demonstrated the potential of RS in distinguishing between normal and cancerous skin tissues [7].

The foundational principle of RS is Raman and Krishnan's groundbreaking study in 1928 which demonstrated the inelastic scattering phenomenon of light in fluids [8], resulting in the Nobel Prize in Physics in 1930. Over the decades, RS has undergone rapid advancements and diversifications, finding applications across various scientific domains, including chemistry and medicine [9]. One of the most promising medical applications of Raman techniques is in the early detection and diagnosis of skin cancer. RS provides a molecular fingerprint of skin tissues by analyzing the vibrational energy of molecules [10]. This non-invasive technique can detect subtle biochemical changes in skin cells that precede the visual signs of skin cancer, offering a potential for early diagnosis. The specificity and sensitivity of RS in identifying malignant tissue changes have made it a valuable tool in cancer diagnostics [11], contributing significantly to early intervention and improved patient outcomes.

In addition to its non-invasive nature, RS signals are also measured regarding Raman frequency shift, calculated using the difference between the peak and excitation laser energy. The unique characteristics make RS signals independent of laser excitation frequency (i.e., the light source), allowing flexible excitation frequency selection depending on characteristics of particular samples and comparison of a spectrum to other spectra even when different laser excitation energies are used [12]. In addition, RS signals are also independent of the optical properties of the tested tissues, i.e., without requiring dyes, nanoparticles, or other contrast agents [13].

Received 7 February 2024; revised 25 May 2024; accepted 28 July 2024. Date of publication 29 August 2024; date of current version 6 December 2024. This work was supported by the U.S. National Science Foundation under Grant IIS-2302786 and Grant IIS-2236579. *(Corresponding author: Xingquan Zhu.)*

Zhiqiang Wang, Yanbin Lin, and Xingquan Zhu are with the Department of Electrical Engineering and Computer Science, Florida Atlantic University, Boca Raton, FL 33431-0992 USA (e-mail: xzhu3@fau.edu).

Andrew C. Terentis is with the Department of Chemistry and Biochemistry, Florida Atlantic University, Boca Raton, FL 33431-0992 USA.

John Strasswimmer is with the College of Medicine, Florida Atlantic University, Boca Raton, FL 33431-0992 USA.

Both code and data used in our study are available at: https://github.com/yeyimilk/tclp.

Digital Object Identifier 10.1109/JBHI.2024.3451950

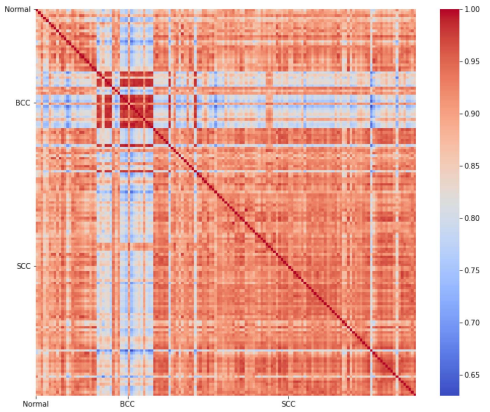

Fig. 1. Heatmap showing intra- and inter-class sample similarities of the RS skin cancer dataset, using original RS signals as features. Samples are indexed based on their class labels. Each value denotes the cosine similarities between a pair of samples. The similarity is color-coded with the transitions from blue, low similarity, to red, high similarity.

A. Learning From RS Data: Challenges

Machine learning (ML) and deep learning (DL) techniques have been widely used in RS data analysis [14], particularly in the medical field, such as detecting skin cancer. Traditional ML algorithms like Linear Discriminant Analysis (LDA), Random Forests (RF), and Support Vector Machines (SVM) have been effectively used in RS data analysis [15], [16]. Moreover, DL has further revolutionized this domain. DL models, especially Convolutional Neural Networks (CNNs) and their variants [17] have demonstrated great proficiency in handling the complexity of spectroscopy data. These advanced models extract complex patterns and features, leading to more accurate and nuanced interpretations. This combination of ML or DL with RS is key in refining diagnostic processes and extracting deeper insights from spectroscopy data.

Nevertheless, the classification of RS data faces significant challenges. Limited data availability, a common issue in medical fields due to privacy concerns, hinders the development of robust ML and DL models [18]. Furthermore, the inherent noise in RS data adds another layer of complexity, potentially deteriorates the classification accuracy [19]. RS data are inherently noisy primarily due to the nature of the Raman scattering process itself, which is intrinsically weak compared to other types of light-matter interactions. The RS signal is generated through inelastic scattering, where a small part of the incident light is scattered at different wavelengths, resulting in a very weak signal often overshadowed by much stronger background noise, such as fluorescence or ambient light. Moreover, collecting and measuring these weak signals require highly sensitive detectors, which can introduce additional noise, such as thermal or electronic noise, complicating the signal-to-noise ratio further [20], [21].

Fig. 1 demonstrates cosine similarity heatmap between samples in an RS skin cancer dataset, using original RS signals as features. It shows that samples have high intra- and inter-class similarities, meaning that samples between classes can exhibit similar or higher similarities than samples within the same class.

The above challenges motivate the development of sophisticated data processing techniques and advanced algorithmic approaches to ensure accurate and reliable classification results. To overcome these challenges, innovative approaches like transfer learning and contrastive learning have been employed in RS data analysis. Transfer learning allows for adapting models pre-trained on extensive datasets to smaller, specific datasets, effectively addressing the issue of limited sample availability [22], [23], [24]. This approach leverages pre-existing knowledge to enhance performance in specific RS classification tasks. On the other hand, contrastive learning can significantly enhance the ability to discern data differences and structural similarities [25], [26], even in challenging scenarios with noise and limited annotated data. This technique effectively narrows the representation gap between similar spectra while distinctly separating dissimilar ones, thereby boosting the accuracy and robustness of spectral data interpretation [27].

B. Contribution and Rationality

Building on these concepts, our study introduces an innovative approach integrating transfer learning with contrastive learning for RS data classification. Specifically, we employ a source dataset to train an encoder utilizing the contrastive learning technique. Subsequently, this trained encoder forms the basis of a classifier, which is further refined and fine-tuned using a target dataset. This methodology leverages the strengths of both transfer and contrastive learning paradigms, aiming to enhance the accuracy and efficiency of RS data classification. Experiments and comparative analyses demonstrate significant improvement in the classification of RS data using our proposed methodology.

The niche of using transfer learning and contrastive learning to alleviate the sample scarcity stems from RS signals' unique characteristics of being independent of excitation laser energy and optical properties of the tested samples [13]. In transfer learning, a source dataset is used to support the learning on the target dataset, where two datasets (i.e., source and target datasets) might be collected using different lasers and different types of tissues. The independent nature of the RS signals, w.r.t lasers, and samples allows us to use source datasets with a similar RS shift to the target dataset to improve the learning and classification performance.

II. Background and Related Work

A. Machine Learning for RS Data Analytics

Combining ML with RS is a promising approach for quick analysis and diagnosis of cancers [15]. Traditional ML models, both linear, such as LDA, or non-linear, such as SVM, Decision Tree (DT), or RF, are usually used with data processing techniques, such as Principal Component Analysis (PCA) for reducing dimensions and capturing the largest variation, and smoothing methods like moving average for despising.

In biomedical RS, the most common ML model is a combination of PCA and LDA [28], where LDA subsequently learns a criterion for classifying the data into one of several categories with labeled examples. RS was obtained from cultured breast cancer cell lines in [16], and the data were analyzed by two ML algorithms: PCA–discriminant function analysis (DFA) and PCA–SVM. Using DT to evaluate the serum tumor markers levels in blood samples, RS was combined with PCA to predict the recurrence of gastric carcinoma (Gca) [29]. DT and RF are

trained to discriminate patients' liver cells between tumor and non-tumor by exploiting a combination with RS [30]. Other feature extraction algorithms: partial least squares (PLS), kernel principal component analysis (KPCA), isometric feature mapping (isomap) and locally linear embedding (LLE), combined with other classifiers, $k-$nearest neighbor (kNN), extreme learning machine (ELM), DT, backpropagation neural network (BP), genetic optimization backpropagation neural network (GA-BP) and LDA, are also able to detect cervical cancer [31].

B. Deep Learning for RS Data Analytics

DL techniques have been widely used as modern tools for accurately classifying and analyzing RS data. Among all DL structures, CNNs, particularly 1D-CNN, have become the most popular leveraged models due to their proficiency in spectral data representation and classification [18]. Studies have demonstrated that 1D-CNN models can achieve outstanding diagnostic accuracies, especially in medical diagnostics, such as in the classification of breast tissues [32], extracellular vesicles, and the differentiation of melanocytes from melanoma cells [33].

Beyond CNNs, other DL structures such as Residual Networks (ResNet) and autoencoders have also been effectively adapted for Raman spectral data analysis to deal with complex problems like outlier robustness and rapid identification across biological samples. ResNet, CNN architectures with residual connections, have shown potential in parsing datasets ranging from human bladder tissues to classifying types of skin tumors, showing substantial superiority over traditional methods [34]. Autoencoders, known for their unsupervised feature learning and anomaly detection capabilities, when combined with locally connected neural networks, have proven to classify the presence of outliers. [35] accurately. Furthermore, Generative Adversarial Networks (GANs) have been utilized ingeniously to enhance data augmentation in improving classification performance even when original datasets are limited [36], [37]. Additionally, recurrent structures, such as recurrent neural networks (RNNs) and Long Short-Term Memory (LSTM) networks, have been proposed for the discrimination of various biological entities with RS data [38].

C. Transfer Learning

It is crucial to overcome dataset size limitations to successfully train classification models in the realm of RS [18]. Recent research highlights the effectiveness of transfer learning in this context. The most popular method involves pre-training a DL model with a domain-specific RS dataset, followed by fine-tuning using source data. This approach effectively utilizes contrastive learning, where a model pre-trained on source data undergoes modification by discarding its original input and output layers while retaining the 1D CNN network with trained weights [22].

Innovations in data input and DL model construction further augment the potential of transfer learning in a decision fusion-based approach [39], which feeds multiple target datasets into diverse DL models like CNN-LSTM, GoogLeNet, and ResNet. The outputs of these models then serve as inputs for logistic regression, which undergoes further refinement with source data for enhanced classification. Meanwhile, other researchers focus on reconstructing input data to facilitate transfer learning. To achieve transfer learning, spline interpolation is used to adjust wavenumber intervals [23], while the performance was surpassed by a polynomial reconstruction algorithm for Raman spectra [24]. These advancements demonstrate that not only the relevance of the source dataset but also the innovative pre-processing and model adaptation significantly boost the accuracy of RS data classification.

D. Contrastive Learning

Recently, contrastive learning methods have attained impressive results in various downstream applications, notably in the field of image recognition [40], [41], [42]. A simple framework for contrastive learning of visual representations (SimCLR) was presented to characterize the similarities and dissimilarities between images in a batch without requiring specialized architectures or a memory bank [25]. Similar samples were encoded as the same feature representations, while dissimilar samples must be encoded as different representations. However, directly applying SimCLR to a time series field usually performed poorly due to unmatched data augmentations and feature extractors. Hence, a framework called TimeCLR was proposed to extend SimCLR to be suitable for univariate time series representation [40], which is achieved by combining the advantages of Dynamic Time Warping (DTW) data augmentation and InceptionTime. Similarly, a self-supervised learning approach was introduced for time-series analysis based on the SimCLR contrastive learning using multiple data-augmentation techniques [41].

Due to the astonishing performance of contrastive learning in image recognition, applying contrastive learning to RS has attracted researchers' interest. Using contrastive learning, the representation learning module can pull similar spectra closer in the learned representation space and simultaneously push dissimilar spectra away [43]. This enables the representation learning module to distinguish subtle differences from RS, even with noise perturbations. Guo et al. [26] developed a method based on contrastive learning for representation learning aimed at deriving a valuable embedding space for spectral data. The embeddings generated through this learning process can not only differentiate spectra from various compounds but also uncover the structural similarities between them. A deep clustering-based framework, RamanCluster, was designed for the accurate and robust unsupervised identification of pathogenic bacteria using Raman spectral analysis, eliminating the requirement for any annotated data. An Xception-based architecture with the Siamese network for Raman spectrum matching was tested in three publicly available datasets, achieving good accuracy in classification [27]. This approach demonstrated commendable robustness in challenging scenarios.

III. Material and Method

A. Datasets

Target dataset $\mathbb{X}_t$*:* Our research aims to study transfer and contrasting learning for skin cancer tissue classification using

TABLE I
STATISTICS OF THE BENCHMARK SOURCE RS DATA

Category	Cells	Serum content	# of samples
A	A2058	10%	9
A-S	A2058	0%	9
G	G361	10%	9
G-S	G361	0%	8
HPM	HPM	10%	9
HPM-S	HPM	0%	9
HF	HF	10%	9
HF-S	HF	0%	9
ZAM	ZAM	10%	9
ZAM-S	ZAM	0%	9
DMEM	None	10%	8
DMEM-S	None	0%	9

TABLE II
STATISTICS OF THE BENCHMARK TARGET RS DATA

Category	# of samples
BCC	36
Normal	63
SCC	50

Raman spectroscopy. The target dataset used in your study was originally collected from Strasswinmer Mobs Surgery, Delray Beach, FL. The same dataset was also used in our previous study about deep learning based RS data classification [36].

Table II reports the statistical breakdown of the benchmark target RS dataset [44], which includes three categories: Basal BCC, Normal, and SCC. It contains 149 records in total, showing an imbalance in distribution. The Normal category has the highest number of records at 63, surpassing SCC with 50 records and significantly exceeding BCC, which has only 36 records. For further information about this dataset, please refer to the study about Raman spectroscopy based SCC and normal skin separation following treatment with a high-powered CO2 laser [44], which investigates the impact of laser treatment on the RS tissue separation.

Source Dataset $\mathbb{X}_s$*:* The source dataset used in our study was originally collected by Erzina et al. (2020) [45], whose objective is to study cancer detection using a combination of SERS (Surface-Enhanced Raman Spectroscopy [46]) and convolutional neural network.

Table I reports the statisitics of the source dataset, which was amassed at the University of Chemistry and Technology, Prague, as part of the research focused on cancer detection. It contains twelve different classes, each containing three unique types of substrate surface RS data. The number of samples in each class is small, and vary between eight to nine samples. Within each sample, there are three data types of benzenediazonium tosylate derivatives were designated: ADT-NH2 (amine-functionalized) (NH2), ADT-COOH (carboxyl-functionalized) (COOH), and ADT-(COOH)2 (dicarboxyl-functionalized) (COOH2), all synthesized according to established procedures detailed in prior research. The Raman shift scale used in this dataset is measured in reciprocal centimeters, ranging from a start point of 100 to an endpoint of 4,278. This range translates into 2,090 data points for each sample.

From Fig. 3, the samples show notable differences in wavenumber lengths between source and target datasets. Specifically, the source data exhibits a longer wavenumber, extending up to 2,090, in contrast to the target data, which spans only 1,608 wavenumbers. However, the latter segment of the source data, particularly in the wavenumber range from 1,608 to 2,090, demonstrates a flattened curve, which is almost a straight line, that bears no resemblance to the target data. Intuitively, this segment will likely have minimal, if not zero, influence on our methodological approach. Consequently, this extra tail segment of the source data has been removed to ensure consistency in the length format of source and target data. This adjustment facilitates a better way of data handling and network training for comparison.

B. Data Augmentation for RS Data

This section delves into the intricacies of data augmentation techniques specifically tailored for RS data, detailing the methodologies and their impact on the original data. For a better understanding, a summary of notations and symbols is presented in Table III.

1) Random Noise: The Random Noise operation introduces a small amount of random noise to each element of the input x. Given a probability p, mean value μ, and standard deviation σ, the operation $N(\cdot)$ adds a random value from the normal distribution $\mathcal{N}(\mu, \sigma^2)$ to each input point with a probability of $1-p$. An example is shown in Fig. 2 under "Noise".

$$N(x_i) = \begin{cases} x_{i,j} + \mathcal{N}(\mu, \sigma^2) \ |\forall j, x_{i,j} \in x_i; & 1-p \\ x_i & p \end{cases} \tag{1}$$

2) Random Scale: The Random Scale operation modifies the input x by scaling up or down its values. Given a probability p and a scaling factor a, the operation $S(\cdot)$ scales the input x by a factor of $r(1-a, 1+a)$ with a probability of $1-p$. An illustration is provided in Fig. 2 under "Scale".

$$v = r(1-a, 1+a) \tag{2}$$

$$S(x_{i,j}) = \begin{cases} v x_{i,j} \ |\forall j, x_{i,j} \in x_i; & 1-p \\ x_i & p \end{cases} \tag{3}$$

3) Random Shift: The Random Shift operation translates the entire input x_i from left to right or right to left. Given a probability p and a maximum shift value a, the operation $S(\cdot)$ shifts the input x_i in the direction $l(b)$, where a positive b means a leftward shift and a negative b stands for a rightward, by $|r_p(-a, a)|$ steps for a probability of $1-p$. An example is presented in Fig. 2 under the title "Shift".

$$v = r_p(-a, a) \tag{4}$$

$$F(x_i) = \begin{cases} l(v)x_i & 1-p \\ x_i & p \end{cases} \tag{5}$$

$$l(v)x_i = x_{i,j} \leftarrow x_{i,k} \quad 0 <= j <= m$$

$$k = ((j+v)\%m + m)\%m \tag{6}$$

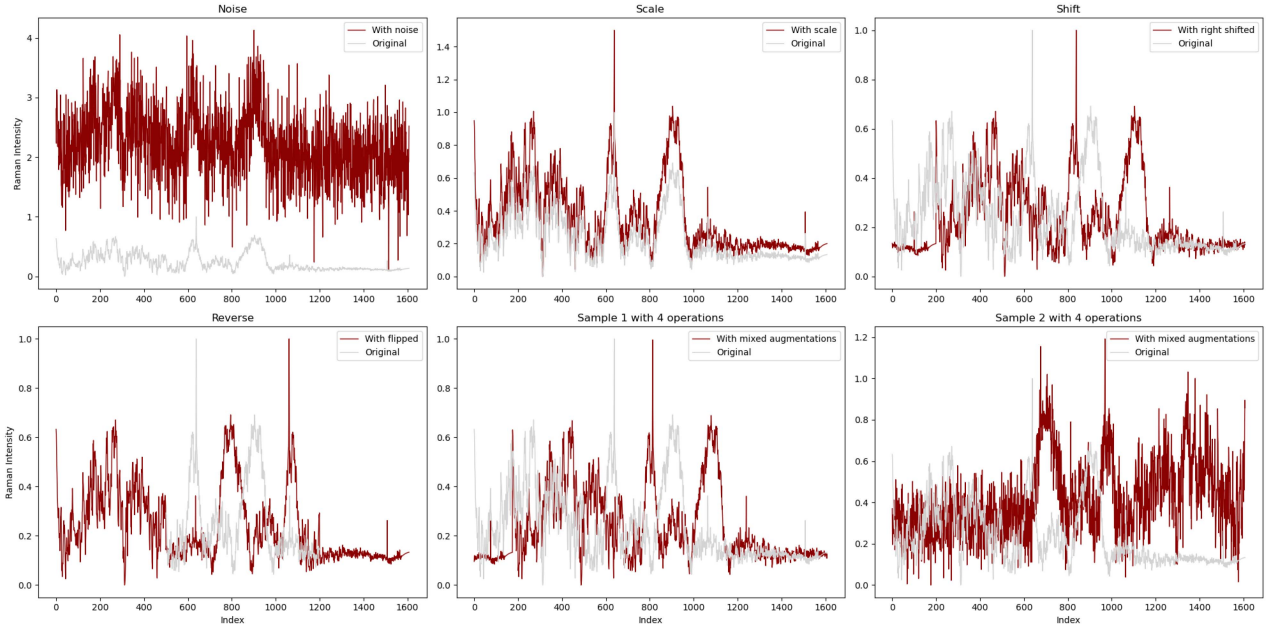

Fig. 2. Examples of RS data augmentation operations and results. From left to right, top to down, each plot represents Random Noise, Random Scale, Random shift, Random Reverse, respectively. The last two are combinations of multiple augmentation operations.

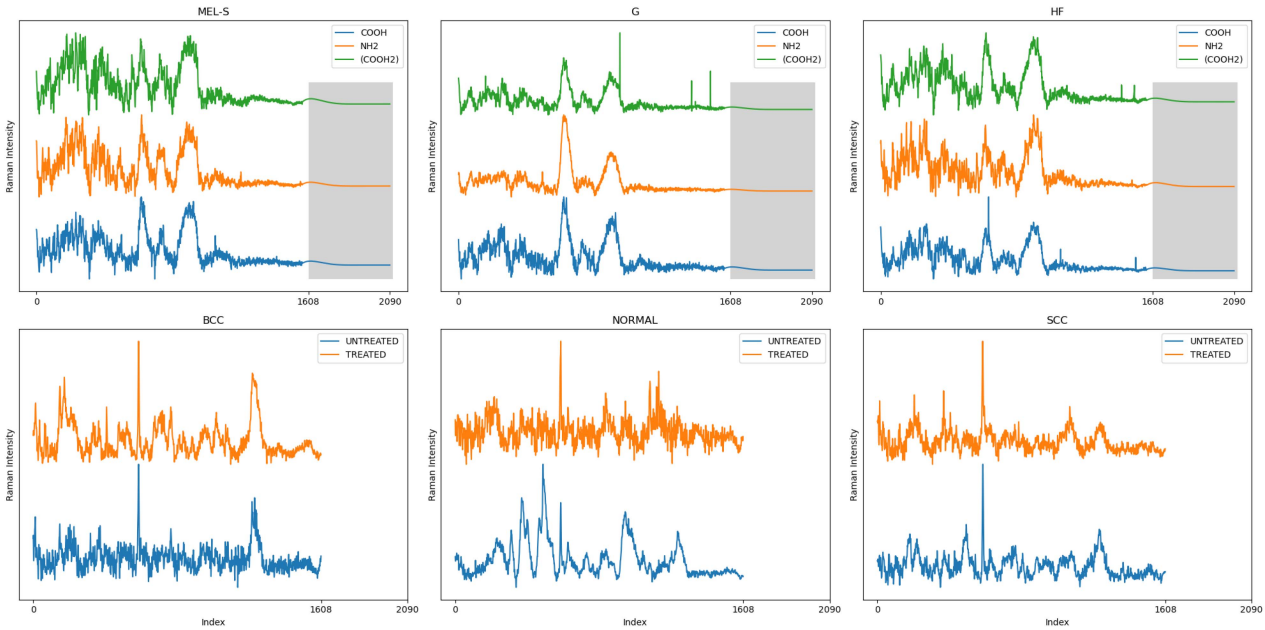

Fig. 3. RS data samples are presented from both the source and target datasets. In the first row, three distinct classes from the source dataset are displayed. It includes twelve types of cancer and normal cells in total. Each class is representative of a specific category within the source data and is characterized by three main spectral bands. These bands arise from the structure of grafted organic moieties, specifically NH_2, COOH, and $COOH_2$ groups. The second row, in contrast, displays three unique classes from the target dataset: BCC, normal, and SCC, respectively. Each sub-figure in this row demonstrates two samples, treated $vs.$ untreated, denoting whether samples have been treated using a high-powered CO2 laser (treated) or not (untreated), respectively.

4) Random Reverse: The Random Reverse operation inverts a segment of the input x_i. In this context, m denotes the number of channels in x_i, and t is the size of the segment to be reversed.

$$R(x_i) = \begin{cases} Ix_i & 1 - p_1 \\ x_i & p_1 \end{cases} \tag{7}$$

$$Ix_i = \begin{cases} x_{m-j} \; |0 <= j <= m & 1 - p_2 \\ \begin{cases} x_{i,j} \; |0 < j, o + t < j < m \\ x_{t+2k-j} \; |o <= j <= t + o \end{cases} & p_2 \end{cases} \tag{8}$$

Given probabilities p_1, p_2, and a specified reversal size t, the $R(\cdot)$ operation reverses the portion of the input x_i with a probability of $1 - p_1$. Within this operation, it reverses a segment of size

TABLE III
SUMMARY OF NOTATIONS AND SYMBOLS

Notations	Descriptions
$\mathbb{X}$	A Raman Spectroscopy dataset
n	Number of samples
N	Batch size
m	Number of Raman wave shift numbers (number of features)
$x_i \in \mathbb{X}$	A Raman sample in $\mathbb{X}$
$x_{i,j} \in x_i$	Raman shift value reading measured at j wavenumber (cm^{-1} per nanometer nm) for sample x_i
$[\mathtt{min}, \mathtt{max}]$	Minimum and maximum Raman shift wave numbers, $\forall i, j, x_i \in \mathbb{X}: \ \mathtt{min} \leq x_{i,j} \leq \mathtt{max}$
$N(\cdot)$	Add random Gaussian noise to a given Raman sample in wave band level
$S(\cdot)$	Multiply a given Raman sample with a factor in wave band level
$F(\cdot)$	Shift a given Raman sample whether from left or right to left in wave band level
$R(\cdot)$	Reverse part of a given Raman sample in wave band level
$r(a, b)$	Pick up a random number within the range of $[a, b)$
$r_p(a, b)$	Pick up a random integer from the range of $[a, b]$ where both a and b are positive integers
$\mathcal{N}(\mu, \sigma^2)$	Normal distribution with mean, u, and standard deviation, σ
$\Theta_{W_1}(\cdot)$	A DL encoder with learnable weights W_1
$\Phi_{W_2}(\cdot)$	A projection head of dense layers with learnable weights W_2
$\Gamma_{W_3}(\cdot)$	A dense classifier layer with learnable weights W_3

$t_1 = r_p(0, t)$, starting from position $o = r_p(0, t_1)$, with given probability of p_2 as shown in (8). Additionally, the entire x_i is reversed, a process termed "flip", (8), with a probability of $1 - p_2$. This flip operation effectively mirrors the input. However, it is important to note that the reversal of x_i, whether partial or complete, only occurs with a probability of $1 - p_1$, reflecting the stochastic nature of this data augmentation process. The subfigure named "Reverse" shows an example of this operation in Fig. 2.

5) Combinations:

$$T(x_i) = R(N(S(F(x_i)))) \tag{9}$$

The series of data augmentation operations applied to the input x_i is concisely encapsulated in (9). This process sequentially involves shifting the input, followed by scaling, adding noise, and finally, applying a reverse operation. Two examples of this comprehensive T operation can be found in Fig. 2, demonstrating its impact on the input data.

C. Proposed Method: Transfer Contrastive Learning for RS Data

SimCLR [25] is a contrastive learning framework that advanced the field of computer vision by extracting useful and invariant representations. Notably, this approach has been adapted to handle 1D time series data tasks, demonstrating strong efficacy in a variety of applications [40], [47].

Leveraging the success of SimCLR, our study employs this framework in a self-supervised learning context aimed at developing a feature extractor for converting RS data into representation vectors. These vectors are designed to retain the inherent structure and feature information of the original data. Similar to SimCLR, our approach involves contrastive learning to maximize the similarity between two augmented views of the same RS sample. This is achieved by applying a contrastive loss function within the latent, named NT-Xent [25].

Previous studies have demonstrated the effectiveness of transfer learning in classification tasks [48], including classifying RS data [18], [22]. This methodology involves training a DL model with target data and then fine-tuning it with source data. The source data may or may not be directly related to the target data, while better performance was observed when using domain-specific information.

Our study proposes a transfer contrasting learning paradigm that combines contrastive learning with transfer learning. The main theme is to use contrastive learning to tackle the noisy nature of the RS data and leverage the strength of transfer learning to tackle the sample scarcity challenge. The deep neural networks are first trained using RS data from the public domain and then are fine-tuned using target data for better classification accuracy.

Fig. 4 indicates our two-stage approach for cancer tissue classification using RS data combining transfer learning and contrastive learning. Our methodology for cancer tissue classification utilizing RS data employs a two-stage approach integrating transfer learning and the SimCLR technique.

1) Pre-Training Stage (Contrastive Learning): During the initial phase, denoted as the pre-training stage, a series of data augmentations, T, is applied to each sample to generate two transformed instances (x_i and x_j). The purpose is to intentionally augment the signals to teach the underlying deep learning model, such that it can focus on the important patterns instead of nontrivial details. A common approach used in contrastive learning is to augment a sample as two copies (by using different augmentations), such that the classifier must learn to separate whether two copies are from the same sample or not, as defined in the (10).

According to the RS data characteristics, the augmentations used in our study include alterations in scale, noise level, shift, and part of reverse, which help enhance the model's ability to learn robust features. The notation N represents the number of randomly selected mini-batch sizes from the input sample, with each sample undergoing a random augmentation operation twice, resulting in a total of $2N$ samples.

The augmented data is then processed through a neural network encoder ($\Theta_{W_1}(\cdot)$), extracting high-dimensional feature representations (h_i and h_j) using a 1D CNN followed by an MLP structure. Subsequently, a projection head refines these

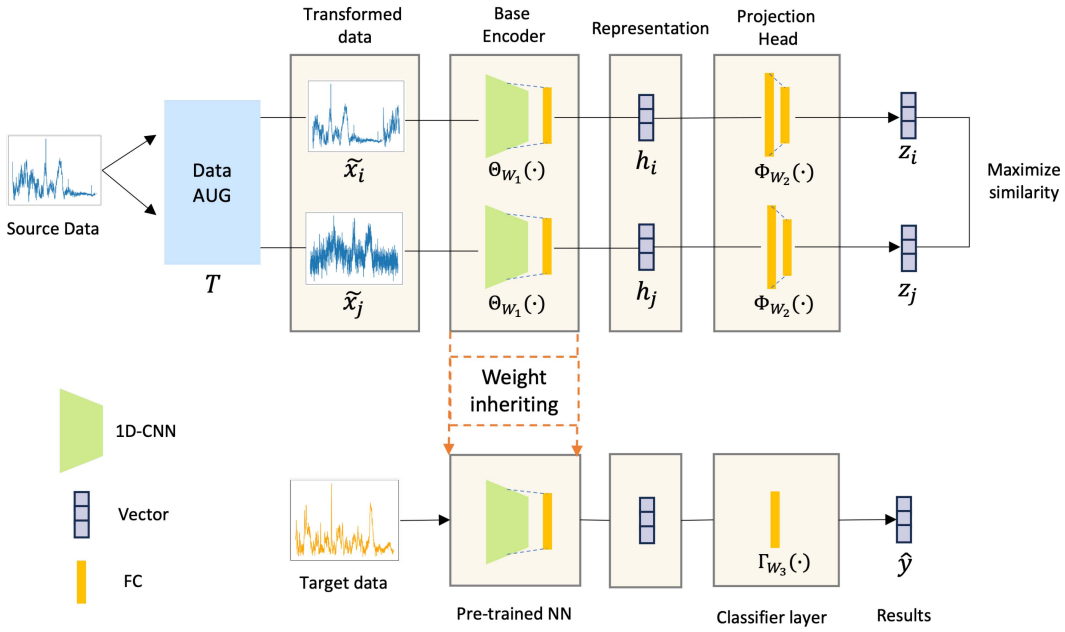

Fig. 4. Top panel: Illustration of the SimCLR working process employed in the pre-training stage of our study. Bottom panel: Description of the subsequent fine-tuning process.

representations, producing (z_i and z_j) through a dense layer followed by a Rectified Linear Unit (ReLU) activation function and another dense layer. This step is significant for distilling essential features and maximizing the representational similarity of augmented pairs, aligning with the core principles of the SimCLR framework.

$$l_{i,j} = -\log \frac{\exp(\mathtt{sim}(z_i, z_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(\mathtt{sim}(z_i, z_k)/\tau)} \tag{10}$$

With the given SimCLR and the defined pre-training stage, the NT-Xent can be defined as shown in (10), where N samples are randomly selected for the contrastive prediction task on pairs of augmented examples, resulting in $2N$ data points. Each augmented pair is defined as a positive pair, while the other $N-1$ pairs serve as negative pairs. The loss function considers each positive pair and computes the final loss across all positive pairs in a mini-batch, N. The similarity between representations of two augmented views of RS data, measured as the dot product of L2-normalized feature vectors (cosine similarity) shown as $sim(\mu, \nu) = \mu^\top \nu / (||\mu|| \cdot ||\nu||)$, should be higher than their similarity with other views. The temperature-scaled (τ) similarities are utilized as logits for cross-entropy. This formula should be calculated for all $i, j \in 2N$.

In this case, assuming a training batch size denoted by N and learnable weights W_1 and W_2 corresponding to the functions $\Theta_{W_1}(\cdot)$ and $\Phi_{W_2}(\cdot)$, respectively, the final objective function for contrastive learning can be defined as presented in (11). This function aims to minimize the associated loss, characterized as follows:

$$\mathcal{L} = \min_{W_1, W_2} \frac{1}{2N} \left[\sum_{k=1}^{N} (l_{2k-1,2k}(x_k) + l_{2k,2k-1}(x_k)) \right] \tag{11}$$

Algorithm 1: TCLP Pre-Training Process.

Input: batch size N, constant τ, structure of Θ_{W_1}, Φ_{W_2}, T, dataset $\mathbb{X}_s$
for *sampled mini-batch* $\{x_k\}_{k=1}^{N} \in \mathbb{X}_s$ **do**
  for *all* $k \in 1, ..., N$ **do**
    $t \sim T, t' \sim T$
    $\tilde{x}_{2k-1} = t(x_k)$
    $h_{2k-1} = \Theta_{W_1}(\tilde{x}_{2k-1})$
    $z_{2k-1} = \Phi_{W_2}(h_{2k-1}, W_2)$
    $\tilde{x}_{2k} = t'(x_k)$
    $h_{2k} = \Theta_{W_1}(\tilde{x}_{2k}, W_1)$
    $z_{2k} = \Phi_{W_2}(h_{2k}, W_2)$
  end
  for *all* $i \in 1, ..., N$, $j \in 1, ..., N$ **do**
    $l_{i,j}$ = Eq. (10)
  end
  $\mathcal{L}$ = Eq. (11)
  update Θ_{W_1} and Φ_{W_2} to minimize $\mathcal{L}$
end
return encoder network Θ_{W_1}.

where the k denotes the sample index within the batch size of N. The terms $2k-1$ and $2k$ correspond to an augmented positive pair derived from sample k. The main idea of this process is shown in Algorithm 1.

2) Fine-Tuning Stage (Transfer Learning): In the second stage, named fine-tuning, we leverage the encoder and representation layer obtained from the first stage, integrating them with a new output layer to construct a specialized classifier. Significantly, the initial weights for the encoder and representation layer are retained from the previous phase. This conservation of weights is essential, as it transfers the classifier to build upon the broad understanding of RS data and signals acquired during pre-training rather than starting from scratch. Doing so ensures

Algorithm 2: TCLP Fine-Tuning Process.

Input: pre-trained Θ_{W_1}, structure of Γ_{W_3}, constant epochs e, dataset $\mathbb{X}_t$, and label $\mathbb{Y}_t$
for $1...e$ **do**
  $\mathbf{h} = f(\mathbb{X}_t)$
  $\hat{y} = \Gamma_{W_3}(\mathbf{h})$
  $\mathcal{L}$ = Eq. (13)
  update Θ_{W_1}, and Γ_{W_3} to minimize $\mathcal{L}$
end
return Θ_{W_1} and Γ_{W_3}

that the sophisticated patterns and features learned in the initial stage are kept and transferred to the specific task of cancer tissue classification.

This classifier is then precisely fine-tuned with a target dataset to ensure the model focuses on the unique characteristics of the target cancer tissue. This targeted approach adjusts the model's learned patterns for more accurate identification and classification of cancer cells. The classification cross-entropy loss function is defined as (13) to be minimized. Where n is the number of samples in the dataset, and C is the number of classes. y_{ij} is the true label (1 for the correct class and 0 for others) and $\hat{y}_{i,j}$ is the predicted probability of the i-th sample belonging to the j-th class, which is determined by the network θ_{W_1} and $\Gamma_{W_3}(\cdot)$. $\hat{\mathbf{y}}_i = (\hat{y}_{i1}, \ldots, \hat{y}_{iC})$.

$$\hat{\mathbf{y}}_i = \Gamma_{W_3}(\Theta_{W_1}(x_i)) \tag{12}$$

$$\mathcal{L} = \min_{W_1, W_3} -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{C} y_{ij}\log(\hat{y}_{i,j}) \tag{13}$$

This stage is critical for transferring the generalized learning from the pre-training stage to the specific task at hand, enhancing the model's accuracy and reliability in real-world diagnostic applications. The algorithm of this stage is presented in Algorithm 2.

IV. Experiment and Results

A. Baselines

1) Traditional Machine Learning Methods: We employ a range of traditional machine learning classifiers as our baseline to compare our model's performance in classifying Raman spectroscopy cancer tissues. These include:

- *Multinomial Naive Bayes (MNB):* A probabilistic classifier known for its simplicity and effectiveness in classification tasks involving multiple classes.
- *Logistic Regression (LR):* A widely-used method for binary classification problems, adapted here for multiclass classification.
- *RF:* An ensemble learning method that combines multiple DTs for improved accuracy and robustness.
- *DT:* A non-parametric approach that divides the dataset into branches to form an inverted DT.
- *kNN:* A simple algorithm that classifies data points based on the majority class of their nearest neighbors.
- *SVM:* A powerful classifier that finds the optimal hyperplane for class separation in a high-dimensional space.

2) Deep Learning Approaches: In addition to traditional methods, we also compare the efficacy of deep learning models in our study:

- *Multilayer Perceptron (MLP):* A basic form of a neural network consisting of multiple layers of perceptrons.
- *CNN-Ma:* We utilized this CNN architecture from Ma et al. (2021), this model offers an approach to handling spectroscopic data for classification [32].
- *ResNet:* This model with ResNet [49] structure was from paper.
- *RNN:* RNN are designed for processing sequences, efficiently capturing temporal dynamics in data [50].
- *LSTM:* A type of RNN capable of learning order dependence in sequence prediction problems [51].
- *Gated Recurrent Units (GRU):* they are a gating mechanism in RNN [52].

3) Proposed Model and Variants for Ablation Study: We denote our proposed method by TCLP. For ablation study purposes, we also introduce four variants by removing contrastive and/or transfer learning components from TCLP to understand the role each module is playing.

- $\texttt{TCLP}_{\neg(\text{T,C})}$ **:** This is a variant of the TCLP, which removes transfer and contrastive learning from the TCLP. This approach directly applies the CNN model, inspired by the architectures from our previous study [36], to the target dataset to train a deep learning model for classification.
- $\texttt{TCLP}_{\neg\text{T}}$ **:** This is a variant of the TCLP, by only incorporating the contrastive learning component (i.e. the transfer learning is removed). Initially, the model was pre-trained on the target dataset using contrastive learning loss, which employs a self-supervised approach to learn rich feature representations. Following this, it undergoes fine-tuning on the same target dataset. This method enhances the model's feature extraction capabilities, potentially leading to a more accurate classification of cancer tissues.
- $\texttt{TCLP}_{\neg\text{C}}$ **:** This is a variant of the TCLP, by only incorporating the transfer learning component (i.e. the contrastive learning is removed). The encoder in the model was pre-trained with source dataset $\mathcal{X}_s$ followed by a fine-tuning stage with target dataset $\mathcal{X}_t$ for the model. When pre-training or fine-tuning the model, no contrastive learning loss is imposed on the learning objective.
- $\texttt{TCLP}$**:** The proposed method utilizes contrastive learning for pre-training on a source dataset. After pre-training, the model is fine-tuned on the target dataset. This approach investigates the effectiveness of transfer learning in improving model performance, especially considering the diversity and complexity of data in Raman spectroscopy cancer tissue analysis.

B. Experimental Setup

The experiment was rigorously designed to ensure the robustness of the results. We use 5-fold cross-validation process and

TABLE IV
CLASSIFICATION RESULTS ON THE TARGET RS DATASET

	with Min-Max Normalization				with PCA ($n = 100$)			
	ACC(↑)	F1(↑)	AUC(↑)	FNR(↓)	ACC(↑)	F1(↑)	AUC(↑)	FNR(↓)
MNB	0.6051	0.5578	0.7812	0.4133	-	-	-	-
LR	0.7455	0.7088	0.9047	0.1471	0.7655	0.7415	0.9068	0.1167
RF	0.6653	0.5960	0.8237	0.1333	0.5977	0.4805	0.7623	0.0000
DT	0.5703	0.5434	0.6636	0.3000	0.4156	0.3846	0.5445	0.5667
KNN	0.6101	0.5901	0.7850	0.4619	0.6168	0.5963	0.7849	0.4476
SVM	0.6253	0.5255	0.8041	0.0667	0.6453	0.5510	0.8304	0.0000
MLP	0.7786	0.7427	0.9218	0.2333	0.7857	0.7679	0.9163	0.1167
RNN	0.7448	0.7088	0.8845	0.2844	0.6244	0.5975	0.7873	0.3743
LSTM	0.6975	0.6742	0.8846	0.2561	0.4763	0.4261	0.6680	0.2867
GRU	0.7584	0.7396	0.9173	0.1250	0.6246	0.5739	0.7400	0.1738
CNN-Ma	0.6094	0.5219	0.8694	**0.0500**	0.4363	0.3068	0.6368	0.1200
TCLP$_{\neg(T,C)}$	0.8126	0.7858	0.9237	0.1452	0.6379	0.6045	0.7849	0.2638
TCLP$_{\neg C}$	0.8124	0.7931	0.9235	0.0952	0.6851	0.6542	0.8286	0.3186
TCLP$_{\neg T}$	0.8260	0.8015	**0.9255**	0.1452	0.5501	0.5139	0.7260	0.2762
TCLP	**0.8329**	**0.8107**	0.9254	0.0786	0.6379	0.5977	0.7641	0.3962

"With PCA" means that PCA was applied to preprocess the data Before learning the models.

early stopping and dropout techniques for each classification method. This approach can help to overcome the randomness inherent in neural network parameter initialization and training, leading to a more accurate evaluation of method performance. Regarding neural network-based methods, despite different architectures, we standardized all parameters, including epochs and augmentation parameters in the contrastive learning process, to facilitate fair comparisons.

A variant of the proposed base model, referred to as TCLP$_{\neg(T,C)}$ in Tables IV is a DL model structured as a 1D-CNN. This model comprises a single CNN layer equipped with 512 filters of kernel size 7. A ReLU activation function follows this layer. Subsequently, the architecture includes a dense layer with 512 units, followed by a ReLU activation. The final component of this model is the output layer.

Additionally, the *MLP* model, showcased in Table IV, features a two-layered hidden structure. The first hidden layer contains 100 units, and the second layer comprises 32 units. Each layer is followed by a ReLU activation function, ensuring non-linear processing capability. Lastly, the *RNN*, *LSTM*, and *GRU* models consist of 100 units for the output space followed by an output layer.

C. Experimental Results

Table IV presents comprehensive performance metrics for standard baseline classifiers and our novel methods for the original data processed using min-max normalization and PCA. The values are averages from five cross-validation results. Our methods outperform all traditional ML and DL approaches in accuracy, F1 score, and false negative rate(FNR) [53]. The FNR was employed because cancer classification is a digital health problem, and the model needs to detect actual cases of skin cancer correctly.

LG achieves the highest performance among the traditional classifiers, recording an accuracy of 0.7455 and 0.7655 when data proceeded with Min-Max normalization PCA ($n = 100$), respectively. While the other traditional methods fall short in effectively classifying the target RS data, with accuracies all below 70%. However, SVM achieved the lowest FNR with only 0.0667.

Regarding DL approaches, the MLP demonstrates noticeable efficacy in classification, achieving accuracies of 0.7784 and 0.7857 when data was handled with normalization and PCA, respectively. The CNN-Ma shows the lowest FNR, 0.05, among all the models, along with low values in another measurement, with only 0.6094 accuracy.

However, our new models, the new base model TCLP$_{\neg(T,C)}$, with contrastive learning, TCLP$_{\neg T}$, with transfer learning, TCLP$_{\neg C}$, and our proposed method, TCLP, outperformed these benchmarks in all ACC, F1, and AUC metrics. TCLP achieved outstanding metrics, with the highest accuracy of 0.8329, F1 score of 0.8107, and a high AUC of 0.9254, indicating a significant improvement over existing deep learning models. Despite TCLP achieving a slightly higher FNR score of 0.0786 compared to CNN-Ma's 0.0500, it demonstrates superior performance in other metrics. TCLP notably improves accuracy by over 20% (from 0.6094 to 0.8329) and F1 score by over 30% (from 0.5219 to 0.8107), showing its effectiveness in the task. This improvement in F1 score is also remarkable.

Relative to the best-performing by MLP model, the improvements in accuracy were 3.4%, 4.74%, and 5.43% for TCLP$_{\neg(T,C)}$, TCLP$_{\neg T}$, and TCLP, respectively. The AUC values of the MLP model are comparable to our methods, suggesting similar overall classification effectiveness. However, our methods demonstrate superior accuracy, F1 scores, and lower FNR, which are crucial for reliable cancer tissue classification. This enhanced performance highlights the effectiveness of the innovative techniques and optimizations integrated into our approaches.

One notable observation is the distinct performance exhibited by the new base model when enhanced through transfer learning, denoted as TCLP$_{\neg C}$. When the data is pre-processed

TABLE V
PERFORMANCE OF TCLP USING DIFFERENT DL ENCODER STRUCTURES

Encoders	Method	ACC(↑)	F1(↑)	AUC(↑)	FNR(↓)
MLP	$\text{TCLP}_{\neg(\text{T,C})}$	0.7786	0.7427	0.9218	0.2333
	$\text{TCLP}_{\neg\text{C}}$	0.7853	0.7533	0.9081	0.2000
	$\text{TCLP}_{\neg\text{T}}$	0.7920	0.7555	0.9211	0.2178
	TCLP	0.7855	0.7466	0.9128	0.1778
CNN-Ma	$\text{TCLP}_{\neg(\text{T,C})}$	0.6094	0.5219	0.8694	0.0500
	$\text{TCLP}_{\neg\text{C}}$	0.6779	0.6459	0.9043	0.0667
	$\text{TCLP}_{\neg\text{T}}$	0.6586	0.5174	0.8635	0.0400
	TCLP	0.7715	0.7356	0.9012	0.1111
ResNet	$\text{TCLP}_{\neg(\text{T,C})}$	0.7515	0.7176	0.9289	0.2250
	$\text{TCLP}_{\neg\text{C}}$	0.7579	0.7086	0.9155	0.3300
	$\text{TCLP}_{\neg\text{T}}$	0.7579	0.7044	0.9226	0.1000
	TCLP	0.7717	0.7341	0.9276	0.2000
RNN(100)	$\text{TCLP}_{\neg(\text{T,C})}$	0.7448	0.7088	0.8845	0.2844
	$\text{TCLP}_{\neg\text{C}}$	0.7055	0.6525	0.8394	0.1167
	$\text{TCLP}_{\neg\text{T}}$	0.7251	0.6985	0.8851	0.2000
	TCLP	0.7113	0.6679	0.8650	0.1844
LSTM(100)	$\text{TCLP}_{\neg(\text{T,C})}$	0.6975	0.6742	0.8846	0.2561
	$\text{TCLP}_{\neg\text{C}}$	0.6453	0.6053	0.7681	0.2390
	$\text{TCLP}_{\neg\text{T}}$	0.7175	0.6967	0.9028	0.2633
	TCLP	0.7515	0.7293	0.8897	0.2619
GRU(100)	$\text{TCLP}_{\neg(\text{T,C})}$	0.7584	0.7396	0.9173	0.1250
	$\text{TCLP}_{\neg\text{C}}$	0.7315	0.6941	0.8796	0.2550
	$\text{TCLP}_{\neg\text{T}}$	0.7851	0.7673	0.9142	0.1488
	TCLP	0.7922	0.7772	0.9173	0.1250

MLP: Dense(100) x Dense(32)
CNN-Ma: CNN(10x10) x MP(2) x BN x Dense(100) x BN
ResNet: ResNet[(CNN(64) x 2)] x 3 x Dense(500)

using Min-Max normalization, both $\text{TCLP}_{\neg\text{T}}$ and TCLP outperform the base model, $\text{TCLP}_{\neg(\text{T,C})}$, for all metrics. However, $\text{TCLP}_{\neg\text{C}}$ negatively impacts the accuracy, which is marginally slower than that in the base model but shows great improvement in the FNR score.

In scenarios where the PCA technique is employed, only 4 out of 14 models show an improvement in accuracy, and 5 out of 14 show an improvement in the FNR. Those models are primarily from traditional ML algorithms and MLP. Table IV indicates that the PCA method does not yield positive results for NN models, except for MLP. For NN models, PCA significantly reduces accuracy and increases the FNR rate. These two metrics are crucial for the cancer classification task. These results suggest that NN models tend to utilize the entirety of the data within each RS data rather than relying solely on a reduced set of principal components extracted through PCA.

D. Performance of TCLP Using Different DL Encoders

To further evaluate the efficacy of our proposed method, TCLP, we compare its performance across six distinct DL encoders. Table V presents the detailed results for these models. Overall, TCLP effectively improves the accuracy for five out of six encoders and reduces the FNR for three encoders, with one showing no change.

When considering only transfer learning ($\text{TCLP}_{\neg\text{C}}$), improvements in accuracy were observed in half of the models, accompanied by a decrease in FNR. In contrast, when solely implementing contrastive learning ($\text{TCLP}_{\neg\text{T}}$), enhancements in accuracy were noted for five out of six models, with four demonstrating a reduction in FNR values. The integration of both techniques under TCLP showcased its efficacy by further elevating accuracy across five models; while its FNR was lower than base model ($\text{TCLP}_{\neg(\text{T,C})}$), none achieved the lowest FNR.

Specifically utilizing a DL as an encoder revealed that all three methods—transfer learning, contrastive learning, and their combination—resulted in diminished accuracy but concurrently exhibited a decrease in FNR.

Using an MLP encoder, the accuracy increase was marginal, but the FNR decreased significantly from 0.2333 to 0.1778. Although the FNR increased from 0.0500 to 0.1111 with the CNN-Ma encoder, accuracy was significantly improved from 0.6094 to 0.7715, approximately a 17% increase. Employing a ResNet encoder increased accuracy (from 0.7515 to 0.7717) and a decrease in FNR (from 0.2250 to 0.2000). Using an RNN encoder, while there was a decrease in accuracy from 0.7448 to 0.7113, the FNR was substantially reduced from 0.2844 to 0.1844. For the LSTM encoder, accuracy increased from 0.6975 to 0.7515, but there was a slight increase in FNR from 0.2561 to 0.2619, which is an undesirable outcome. For a GRU encoder, accuracy improved from 0.7584 to 0.7922, while the FNR remained stable at 0.1250.

In comparing the metrics of accuracy and FNR between the baseline model based on different encoders and with the proposed method, TCLP, and we found that although the method does not uniformly enhance performance across all NN encoders, it generally leads to significant improvements in both accuracy and FNR for this cancer classification task. This demonstrates the potential utility of our method in enhancing model performance in specific applications.

TABLE VI
CONFUSION MATRIX FOR CANCER CLASSIFICATION

Actual/Predicted	BCC	NORMAL	SCC
BCC	23	7	6
NORMAL	3	55	5
SCC	1	3	46

TABLE VII
PERFORMANCE METRICS FOR CANCER CLASSIFICATION

Class	Precision	Recall	F1 Score	FNR
BCC	0.852	0.639	0.727	0.361
Normal	0.846	0.873	0.859	0.127
SCC	0.807	0.920	0.860	0.080

E. Comprehensive Results

Table IV reports the measurements based on averaging of the five-fold cross-validation results. In Table VI, we report the confusion matrix with respect to each class of the target dataset (using five-fold cross-validations). Table VII further reports the classification performance using different metrics, based on the results reported in the Table VI. Both Tables VI and VII represent the combined performance of the model across all cross-validations, offering a more comprehensive assessment of the model performance.

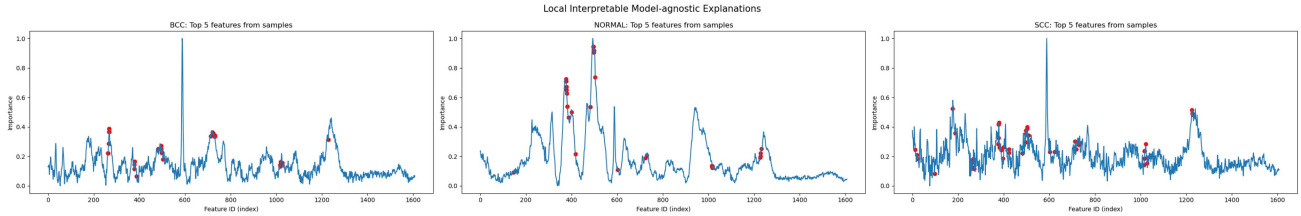

Fig. 5. Aggregated top-5 RS features selected by the LIME explainer [54] for all test samples. Each dot denotes one identified RS feature. Each plot shows the results of all test samples in each individual target class. LIME explainer selects top-5 RS features for each test sample, so each plot has more than five red dots.

Table VI demonstrates that the model prefers classifying data into the Normal and SCC classes. It incorrectly classified 10 and 11 data points as Normal and BCC, respectively, while only 4 data points were incorrectly classified as BCC. It also shows that more data in the BCC and Normal could not be classified correctly. This could be due to the similarities in the features, which can be further proved by the heat map shown in Fig. 1 that a small part of BCC and Normal data are highly similar while differing from other data points.

In Table VII, it is evident that while the precision of the BCC class is notably high at 0.852, its recall (0.639) and F1 score (0.727) are significantly lower, resulting in a higher FNR of 0.361. This indicates that the model struggles to classify BCC tissue correctly and effectively. On the other hand, the SCC class demonstrates a high recall value of 0.920 and a low FNR of 0.080, suggesting that the model performs well in classifying SCC data. However, it is important to note that other data types are more likely to be misclassified into this class, as indicated by its relatively low precision value of 0.807.

F. Model Performance Vs. Feature Explainability

To better explain the model, the LIME explainer [54] was used to find top-5 important features for each of the test samples, shown in Fig. 5. The results show that each sample and each class has its own unique RS features for classification. Peak RS signals do not necessarily form explainable features for separation. Based on the top features, a new dataset was created from the target dataset and then used to validate the model again. However, training and validating models on this new dataset led to a significant decrease in performance metrics. This suggests that DL models may prefer learning from all features rather than specific ones, a hypothesis supported by the PCA results in Table IV, where most NN models exhibit poorer performance after the PCA process.

G. Model Convergence and Overfitting

Fig. 6 illustrates the training and testing losses across epochs. In the NN structure, a dropout rate of 0.2 was implemented, and early stopping with patience of 10 epochs based on the test loss was employed to prevent over-fitting during training. As shown in Fig. 6, without early stopping, the training loss steadily decreases, but the testing loss starts to increase after the early stopping point, indicating the onset of over-fitting. The higher

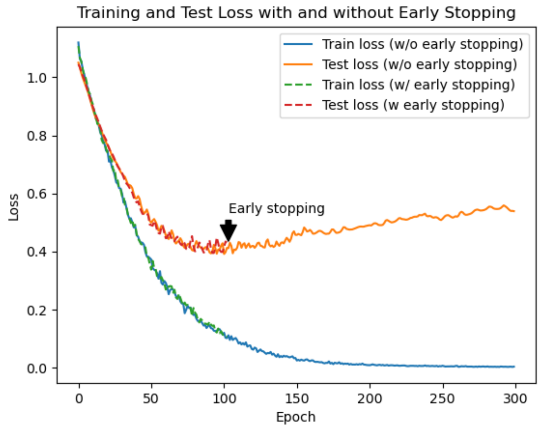

Fig. 6. Training *vs.* test loss with and without early stopping technique to help tackle overfitting.

test loss compared to the training loss is possibly due to noise or outliers in the dataset.

H. Discussion

Both transfer learning and contrastive learning play important roles in TCLP. In reality, selecting source data to support learning on the target data is nontrivial, and we have carried out following three criteria in our design to select domains for RS data. (1) Domain Similarity: We aimed to select domains closely related to cancer or tissue-related Raman spectroscopy data, as these domains are more likely to share similar patterns with our target data. The closer the similarity, the better the transfer learning performance is expected to be; (2) Pattern Similarity: The RS data from the selected domains should exhibit patterns that are not significantly different from our target data. The more similar the patterns, the more effective the transfer learning is expected to be; (3) Effectiveness of Transfer Learning: While the domains we found were not perfectly similar to our target domain, they still demonstrated effectiveness for transfer learning. Finding a better and larger public dataset that closely matches the target domain could lead to better results.

V. Conclusion and Future Work

In this paper, we proposed a transfer contrastive learning paradigm for skin cancer tissue classification using Raman spectroscopy data. We argued that the non-invasive optical nature and

other characteristics of the RS signals, such as the independence of laser sources and optical properties of the tested tissues, make them useful for clinical usage. Nevertheless, RS data are inherently noisy, and the scarcity of samples in biomedical domains further imposes significant challenges to training deep learning models from RS data. By leveraging contrastive learning to handle noisy data and using transfer learning to tackle the sample scarcity, TCLP method can further improve existing DL structures, highlighting the significant potential of our approach in advancing the classification of cancer tissues using RS data.

Looking ahead, our research opens avenues for further exploration into applying advanced DL techniques to medical imaging and diagnostics, especially in scenarios constrained by data privacy and availability. Future work could focus on optimizing these techniques for other types of medical data by generalizing the workflow and models and exploring the implications of these methods in clinical settings. Additionally, investigating our approach's integration with other medical data modalities could provide a more holistic view of medical diagnostics, potentially leading to more accurate and early detection of various cancer types.

Our study also shows some limitations. While our method is effective for most DL structures, it may not be suitable for certain models like RNNs. Furthermore, while our approach improves accuracy significantly, it may also adversely affect the FNR score. Additionally, our dataset is limited to 149 samples, some of which may contain noise from non-standardized manual operations or machine biases. Future work could focus on standardizing the data collection process and expanding the dataset to address these limitations.

REFERENCES

[1] R. S. Stern, "Prevalence of a history of skin cancer in 2007: Results of an incidence-based model," *Arch. Dermatol.*, vol. 146, no. 3, pp. 279–282, 2010.

[2] "Burden of skin disease," 2017, Accessed: Dec. 14, 2023. [Online]. Available: https://www.aad.org/member/clinical-quality/clinical-care/bsd

[3] "Skin cancer," 2022, Accessed: Dec. 14, 2023. [Online]. Available: https://www.aad.org/media/stats-skin-cancer

[4] J. Ferlay et al., "Cancer statistics for the year 2020: An overview," *Int. J. Cancer*, vol. 149, no. 4, pp. 778–789, 2021.

[5] U. Leiter, U. Keim, and C. Garbe, "Epidemiology of skin cancer: Update 2019," in *Sunlight, Vitamin D and Skin Cancer*. Berlin, Germany: Springer, 2020, pp. 123–139.

[6] A. N. Kauvar, T. Cronin Jr, R. Roenigk, G. Hruza, and R. Bennett, "Consensus for nonmelanoma skin cancer treatment: Basal cell carcinoma, including a cost analysis of treatment methods," *Dermatologic Surg.*, vol. 41, no. 5, pp. 550–571, 2015.

[7] J. Zhang, Y. Fan, Y. Song, and J. Xu, "Accuracy of Raman spectroscopy for differentiating skin cancer from normal tissue," *Medicine*, vol. 97, no. 34, 2018.

[8] J. R. Ferraro, *Introductory Raman Spectroscopy*. Amsterdam, The Netherlands: Elsevier, 2003.

[9] R. S. Das and Y. Agrawal, "Raman spectroscopy: Recent advancements, techniques and applications," *Vibrational Spectrosc.*, vol. 57, no. 2, pp. 163–176, 2011.

[10] N. Colthup, *Introduction to Infrared and Raman Spectroscopy*. Amsterdam, The Netherlands: Elsevier, 2012.

[11] X. Feng et al., "Biophysical basis of skin cancer margin assessment using Raman spectroscopy," *Biomed. Opt. Exp.*, vol. 10, no. 1, pp. 104–118, 2019.

[12] R. J. McNichols and G. L. Cote, "Optical glucose sensing in biological fluids: An overview," *J. Biomed. Opt.*, vol. 5, no. 1, pp. 5–16, 2000, doi: 10.1117/1.429962.

[13] L. Xi and H. Jiang, "Image-guided surgery using multimodality strategy and molecular probes," *WIREs Nanomedicine Nanobiotechnol.*, vol. 8, no. 1, pp. 46–60, 2016.

[14] S. Li, J. O. Nyagilo, D. P. Dave, and J. Gao, "Models and methods for quantitative analysis of surface-enhanced Raman spectra," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 2, pp. 525–536, Mar. 2014.

[15] L.-W. Shang et al., "Fluorescence imaging and Raman spectroscopy applied for the accurate diagnosis of breast cancer with deep learning algorithms," *Biomed. Opt. Exp.*, vol. 11, no. 7, pp. 3673–3683, 2020.

[16] L. Zhang et al., "Raman spectroscopy and machine learning for the classification of breast cancers," *Spectrochimica Acta Part A, Mol. Biomol. Spectrosc.*, vol. 264, 2022, Art. no. 120300.

[17] M. Kazemzadeh, C. L. Hisey, K. Zargar-Shoshtari, W. Xu, and N. G. Broderick, "Deep convolutional neural networks as a unified solution for Raman spectroscopy-based classification in biomedical applications," *Opt. Commun.*, vol. 510, 2022, Art. no. 127977.

[18] R. Luo, J. Popp, and T. Bocklitz, "Deep learning for Raman spectroscopy: A review," *Analytica*, vol. 3, no. 3, pp. 287–301, 2022.

[19] M. J. Baker et al., "Clinical applications of infrared and Raman spectroscopy: State of play and future challenges," *Analyst*, vol. 143, no. 8, pp. 1735–1757, 2018.

[20] J. Smulko, M. S. Wróbel, and I. Barman, "Noise in biological Raman spectroscopy," in *2015 Int. Conf. Noise Fluctuations*, 2015, pp. 1–6.

[21] P. M. Ramos and I. Ruisánchez, "Noise and background removal in Raman spectra of ancient pigments using wavelet transform," *J. Raman Spectrosc.*, vol. 36, no. 9, pp. 848–856, 2005.

[22] J. Hu et al., "Raman spectrum classification based on transfer learning by a convolutional neural network: Application to pesticide detection," *Spectrochimica Acta Part A, Mol. Biomol. Spectrosc.*, vol. 265, 2022, Art. no. 120366.

[23] R. Zhang et al., "Transfer-learning-based Raman spectra identification," *J. Raman Spectrosc.*, vol. 51, no. 1, pp. 176–186, 2020.

[24] L.-W. Shang et al., "A novel polynomial reconstruction algorithm-based 1D convolutional neural network used for transfer learning in Raman spectroscopy application," *J. Raman Spectrosc.*, vol. 53, no. 2, pp. 237–246, 2022.

[25] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 1597–1607.

[26] H. Guo, K. Xue, H. Sun, W. Jiang, and S. Pu, "Contrastive learning-based embedder for the representation of tandem mass spectra," *Anal. Chem.*, vol. 95, no. 20, pp. 7888–7896, 2023.

[27] B. Li, M. N. Schmidt, and T. S. Alstrøm, "Raman spectrum matching with contrastive representation learning," *Analyst*, vol. 147, no. 10, pp. 2238–2246, 2022.

[28] F. Picot et al., "Data consistency and classification model transferability across biomedical Raman spectroscopy systems," *Trans. Biophotonics*, vol. 3, no. 1, 2021, Art. no. e202000019.

[29] Z. Guleken et al., "An application of Raman spectroscopy in combination with machine learning to determine gastric cancer spectroscopy marker," *Comput. Methods Programs Biomed.*, vol. 234, 2023, Art. no. 107523.

[30] L. Aversano et al., "Using machine learning for classification of cancer cells from Raman spectroscopy," in *DeLTA*, 2022, pp. 15–24.

[31] M. Li et al., "A novel and rapid serum detection technology for non-invasive screening of gastric cancer based on Raman spectroscopy combined with different machine learning methods," *Front. Oncol.*, vol. 11, 2021, Art. no. 665176.

[32] D. Ma, L. Shang, J. Tang, Y. Bao, J. Fu, and J. Yin, "Classifying breast cancer tissue by Raman spectroscopy with one-dimensional convolutional neural network," *Spectrochimica Acta Part A, Mol. Biomol. Spectrosc.*, vol. 256, 2021, Art. no. 119732.

[33] X. Qiu et al., "Raman spectroscopy combined with deep learning for rapid detection of melanoma at the single cell level," *Spectrochimica Acta Part A, Mol. Biomol. Spectrosc.*, vol. 286, 2023, Art. no. 122029.

[34] I. A. Bratchenko, L. A. Bratchenko, Y. A. Khristoforova, A. A. Moryatov, S. V. Kozlov, and V. P. Zakharov, "Classification of skin cancer using convolutional neural networks analysis of Raman spectra," *Comput. Methods Programs Biomed.*, vol. 219, 2022, Art. no. 106755.

[35] C.-S. Ho et al., "Rapid identification of pathogenic bacteria using Raman spectroscopy and deep learning," *Nature Commun.*, vol. 10, no. 1, p. 4927, 2019.

[36] M. Wu, S. Wang, S. Pan, A. C. Terentis, J. Strasswimmer, and X. Zhu, "Deep learning data augmentation for Raman spectroscopy cancer tissue classification," *Sci. Rep.*, vol. 11, no. 1, p. 23842, 2021.

[37] B. Liu et al., "Laser tweezers Raman spectroscopy combined with deep learning to classify marine bacteria," *Talanta*, vol. 244, 2022, Art. no. 123383.

[38] P. Wang et al., "Discrimination of blood species using Raman spectroscopy combined with a recurrent neural network," *OSA Continuum*, vol. 4, no. 2, pp. 672–687, 2021.

[39] C. Chen et al., "A new method for Raman spectral analysis: Decision fusion-based transfer learning model," *J. Raman Spectrosc.*, vol. 54, no. 3, pp. 314–323, 2023.

[40] X. Yang, Z. Zhang, and R. Cui, "TIMECLR: A self-supervised contrastive learning framework for univariate time series representation," *Knowl.-Based Syst.*, vol. 245, 2022, Art. no. 108606.

[41] J. Pöppelbaum, G. S. Chadha, and A. Schwung, "Contrastive learning based self-supervised time-series analysis," *Appl. Soft Comput.*, vol. 117, 2022, Art. no. 108397.

[42] Y. Liu, Q. Yan, and A. Alahi, "Social NCE: Contrastive learning of socially-aware motion representations," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 15118–15129.

[43] J. Xu, X. Li, Z. Guo, W. E. Huang, and J.-X. Cheng, "Fingerprinting bacterial metabolic response to erythromycin by Raman-integrated mid-infrared photothermal microscopy," *Anal. Chem.*, vol. 92, no. 21, pp. 14459–14465, 2020.

[44] S. A. Fox, A. A. Shanblatt, H. Beckman, J. Strasswimmer, and A. C. Terentis, "Raman spectroscopy differentiates squamous cell carcinoma (SCC) from normal skin following treatment with a high-powered CO_2 laser," *Lasers Surg. Med.*, vol. 46, no. 10, pp. 757–772, 2014.

[45] M. Erzina et al., "Precise cancer detection via the combination of functionalized sers surfaces and convolutional neural network with independent inputs," *Sensors Actuators B, Chem.*, vol. 308, 2020, Art. no. 127660.

[46] X. X. Han, R. S. Rodriguez, C. L. Haynes, Y. Ozaki, and B. Zhao, "Surface-enhanced Raman spectroscopy," *Nature Rev. Methods Primers*, vol. 1, no. 87, 2021.

[47] M. N. Mohsenvand, M. R. Izadi, and P. Maes, "Contrastive representation learning for electroencephalogram classification," in *Proc. Mach. Learn. Health.*, 2020, pp. 238–253.

[48] V. Cheplygina, I. P. Pena, J. H. Pedersen, D. A. Lynch, L. Sørensen, and M. De Bruijne, "Transfer learning for multicenter classification of chronic obstructive pulmonary disease," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 5, pp. 1486–1496, Sep. 2018.

[49] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[50] J. L. Elman, "Finding structure in time," *Cogn. Sci.*, vol. 14, no. 2, pp. 179–211, 1990.

[51] A. Graves and A. Graves, "Long short-term memory," in *Supervised Sequence Labelling with Recurrent Neural Networks*. Berlin, Germany: Springer, 2012, pp. 37–45.

[52] K. Cho et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Proc. Conf. Emp. Methods Natural Lang. Process.*, 2014, pp. 1724–1734.

[53] L. Di Biasi, F. De Marco, A. Auriemma Citarella, M. Castrillón-Santana, P. Barra, and G. Tortora, "Refactoring and performance analysis of the main CNN architectures: Using false negative rate minimization to solve the clinical images melanoma detection problem," *BMC Bioinf.*, vol. 24, no. 1, p. 386, 2023.

[54] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should i trust you?' explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2016, pp. 1135–1144.