Transcending Two-Path Impedance Spectroscopy with Machine Learning:
A Computational Study on Modeling and Quantifying Electric Bipolarity of Epithelia

Benjamin Schindler

Universität Leipzig
Machine Learning Group
Leipzig, Germany
bschindler@informatik.uni-leipzig.de

Dorothee Günzel

Charité - Universitätsmedizin Berlin
Institute of Clinical Physiology
Berlin, Germany
dorothee.guenzel@charite.de

Thomas Schmid

Lancaster University in Leipzig /
Universität Leipzig
Leipzig, Germany
schmid@informatik.uni-leipzig.de

Abstract—**Quantifying tissue permeability is a central task in assessing pathophysiology of intestinal epithelia. A common and convenient approach for this task is to determine electric properties like resistance and capacitance of the epithelial tissue by applying impedance spectroscopy. While the measurement technique itself is well-established, analysis tools and strategies are still subject to ongoing research in epithelial physiology. Estimations of electric parameters are known to be particularly imprecise for models where apical and basolateral sides of the tissue differ significantly from each other. One-sided application of substances such as Nystatin play an important role here, as they alter membrane conductivity on one side of the tissue while leaving other properties unchanged. Here, we present a novel method that considers two functional states of the cells, namely before and after apical addition of the substance Nystatin. To this end, an extensive dataset modeled after the epithelial cell lines HT-29/B6, IPEC-J2, and MDCK I was synthesized. In a broad study, we show that considering features from two distinct tissue states leads to significantly better regressions by decision trees, random forests, and multilayer perceptrons. Therein, we extend previous work in order to progress from a two-path to a more revealing three-path model of electric tissue properties. Parameters of a corresponding equivalent circuit could be determined with less than five percent deviation from the known target value on average. In a post-processing step, predictions by independent machine learning regressions are employed to initialize a least squares parameter fitting, where the associated impedance spectrum is aligned with the originally observed spectrum, reducing the residual sum of squares by 99% on average.**

Keywords–*Physiology; Epithelia; Impedance Spectroscopy; Machine Learning; Least Squares; Neural Networks; Random Forests.*

I. Introduction

Epithelia play a crucial role in animal and human bodies. They are tissues that form barriers between different compartments of an organism as well as between the organism and the environment. Stratified epithelia, such as the skin, consist of multiple cell layers, whereas simple epithelia (e.g., intestinal, kidney, lung, or glandular epithelia) consist of a single cell layer that is attached to the basal lamina, a matrix of extracellular proteins. The results presented here extend previous work aiming at modeling electric properties of these simple epithelia [1] and reflects properties of animal and human colon, intestinal and kidney tissue.

Fully differentiated epithelial cells are polarized, i.e., the basolateral plasma membrane attached to the basal lamina or facing the neighboring cells functionally differs from the opposite, apical plasma membrane (facing the external environment or the lumen of body cavities). Both, apical and basolateral membranes contain a multitude of different ion channels, carriers and pumps (for review see, e.g., [2][3]) and their apical/basolateral compartmentalization is a prerequisite for vectorial transepithelial transport. Apical and basolateral membranes are separated by the tight junction (TJ). TJs are a meshwork of protein strands that consist of integral membrane proteins. These proteins bind to their counterparts within the same plasma membrane as well as within the plasma membrane of the neighboring cells and thus serve two functions:

1) *Fence function.* TJs act as a diffusion barrier for proteins within the plasma membrane and prevent membrane proteins from diffusing from the apical to the basolateral membrane compartment.
2) *Gate function.* TJs limit and regulate the passage of water and solutes between cells, i.e., along the paracellular pathway [4].

Since their discovery in the 1990s, members of the TJ-associated MARVEL proteins (TAMP), including, e.g., occludin or tricellulin [5][6] and the claudin protein family [7] have been recognized to be the main constituents of this paracellular barrier. Claudins are of special structural relevance as they are able to spontaneously assemble into a meshwork of TJ-like strands at the contact areas with neighboring, even when overexpressed in non-epithelial cells [8]. As a consequence, many members of the claudin protein family strengthen the barrier function of the TJ, as they seal off the paracellular cleft. However, some claudins (e.g., claudin-2, -10, -15, -17) were found to act as paracellular cation or anion channels within these TJ strands [9].

The barrier strength of an epithelium can be quantified by measuring the transepithelial resistance (TER). TER is the sum of the resistances of the subepithelial connective tissues (subepithelial resistance, R^{sub}) and the epithelial resistance (R^{epi}). The latter, again, consists of two major resistances: the resistance of the transcellular pathway (R^{trans}) and the resistance of the paracellular pathway (R^{para}), which are connected in parallel. Thus, alterations in TER may reflect alterations in any of these three components:

$$TER = R^{sub} + R^{epi} = R^{sub} + \frac{R^{trans} \cdot R^{para}}{R^{trans} + R^{para}} \qquad (1)$$
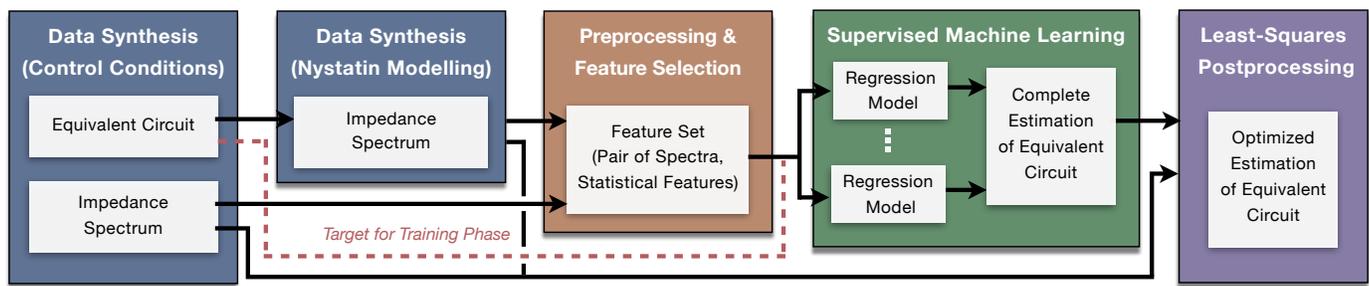
Figure 1. Overview of the presented approach for pairwise analysis of epithelial impedance spectra. The main machine learning tasks (preprocessing, feature selection, supervised learning) are preceeded by data synthesis and followed by a least squares-based postprocessing.

Activation or inactivation of ion channels in the apical or basolateral membrane alters the corresponding membrane conductivities or membrane resistances R^{ap} and R^{bl}, respectively. Any change in R^{ap} and R^{bl} also affects R^{trans} and thereby ultimately alters TER. A prominent example is the over-stimulation of cAMP-regulated Cl^- channels in the apical membrane of enterocytes, e.g., through activation of the adenylate cyclase by toxins such as cholera toxin, toxins from enterotoxigenic *Escherichia coli* (ETEC) or by substances such as caffeine or theophylline. The resulting secretion of Cl^- into the gut lumen is accompanied by osmotic water flux and hence leads to secretory diarrhea.

Alterations in the paracellular barrier of the intestine may be caused by an upregulation of channel-forming TJ proteins or a downregulation of barrier-forming TJ proteins. Both result in a back-leakage of electrolytes and consequently of water into the gut lumen and thus to leak-flux diarrhea (leaky gut syndrome). Dysregulation of TJ proteins is observed in inflammatory diseases, such as chronic inflammatory bowel diseases (Crohn's disease; ulcerative colitis), immune-related diseases (e.g., celiac disease) or in protozoal, bacterial or viral infection (for review see, e.g., [10][11]). Thus, a hall-mark of secretory as well as of leak flux diarrhea is a reduction in TER. Therefore, alterations in TER of a tissue may be due to alterations in the transcellular or the paracellular pathway (or both) and a detailed knowledge of the underlying processes is crucial for an understanding and treatment of these diseases.

Alterations in R^{sub} are observed in many inflammatory diseases that cause increased proliferation of subepithelial cells and thus a thickening of the subepithelial tissue. The resulting increase in R^{sub} is of experimental, not of physiological relevance: in the experiment, TER is increased, suggesting an increase in barrier function. In situ, however, capillaries traverse the subepithelium so that the distance between the capillaries and the basolateral membrane of the epithelium remains unaltered. Thus, the observed increase in TER does not affect the barrier function of the epithelium. Exact knowledge of R^{sub} is thus necessary to correctly assess alterations in R^{epi}.

Many different cell culture models are established to study regulatory aspects of epithelial barrier functions and to investigate the effects of over-expression or knockdown of TJ components. Among the most widely studied epithelial cell lines are human colonic Caco-2, T84, HT-29/B6; human bronchial 16HBE14o-; porcine jejunum IPEC-J2; porcine kidney LLC-PK1; canine kidney MDCK; mouse kidney IMCD-3; mouse

mammary HC11. These cell lines fully differentiate into tight cell monolayers when grown on filter supports, as judged by their transepithelial resistances of several hundreds or even thousands of $\Omega \cdot cm^2$. As in tissue, TER is the sum of R^{epi} and R^{sub}. While R^{epi}, again, consists of R^{trans} and R^{para}, R^{sub} in the cell culture system is the resistance of the filter support on which the cells are grown and thus constant.

In order to describe, explain and understand epithelial functionality, a variety of computational models has been developed for this tissue. To this end, we have established a machine learning-based approach for quantifying electrical properties of epithelia. In particular, we could not only improve estimates for R^{epi} [1], but also the one for the epithelial capacity C^{epi} [12]. Based on an early proof-of-concept study [13], we here introduce an improved, systematic and reliable approach to quantify transcellular properties discerned into apical and basolateral resistances (R^{ap}, R^{bl}) and capacitances (C^{ap}, C^{bl}). The key concept to achieve this is to model machine learning training data, not for individual impedance measurements but a combination of two measurements during distinct tissues states. Moreover, we introduce an additional post-processing step to further improve parameter quantifications. Figure 1 shows an overview of the approach.

The rest of the paper is organized as follows. First, the current state of research is presented (Section II). In the following sections, a new method for the determination of physiological properties of epithelia is introduced. Figure 1 shows this method as a block diagram. The basis is a data synthesis where impedance spectra are synthesized using an equivalent circuit as a cell model (Section III). From previous estimates of physiological properties, epithelial tissues are modeled under control conditions as well as after the addition of Nystatin. From these, two impedance spectra are generated for each sample. In a preprocessing step, new statistical features are constructed from this pair of spectra. Supervised learning is then used to train the prediction of circuit parameters (Section IV). In this step, only a single parameter of the cell model is determined at a time. By combining the individual parameter values, a complete prediction of the model parameters is obtained. Using the originally observed spectra from the data set, the circuit parameters are further optimized in a least squares post-processing (Section V). In a comprehensive study, this method is applied to the three cell lines HT-29/B6, IPEC-J2 and MDCK I. Subsequently, the results of the evaluation are discussed (Section VI) and summarized (Section VII).
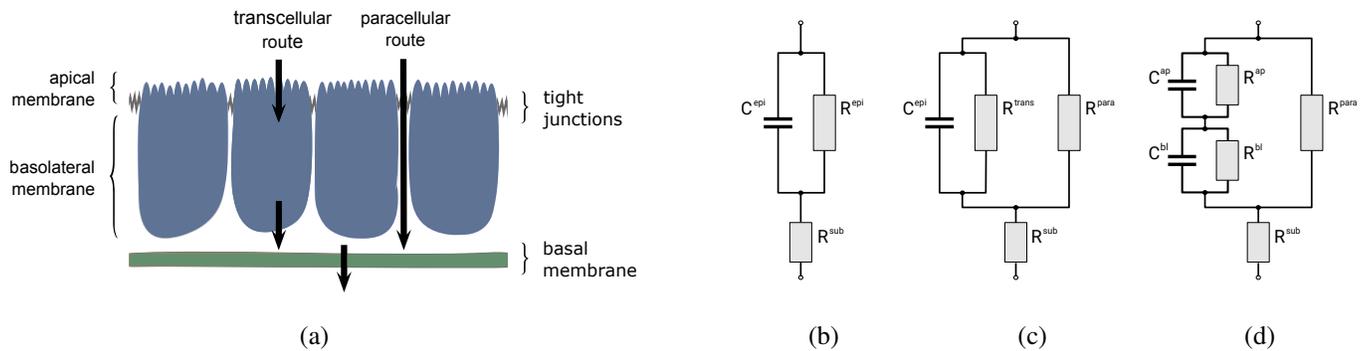
Figure 2. a) Schematic drawing of a simple epithelium. b) A resistor-capacitor (RC) circuit reflects general properties of the epithelium. c) A RC circuit with a parallel resistor reflects the paracellular pathway. d) A 2-RC circuit reflects apical and basolateral membrane properties.

II. State of the Art

Over the last century, a variety of techniques has been established to assess the functionality of epithelia. The most common way to determine tissue permeability is a direct measurement of fluxes, e.g., by using radioactive isotopes or labeled substances. Alternatively, the permeabilities of the two major extracellular ion species, Na^+ and Cl^-, are determined by measuring tissue conductance or its reciprocal, resistance, e.g., by using "chopstick electrodes" or Ussing chambers [14]. Both techniques, however, simply measure TER and are not able to discriminate between any partial resistances (cf. Eq. (1)).

In contrast to DC resistance measurements, impedance spectroscopy measurements allow us to distinguish between R^{sub} and R^{epi}. Further, impedances reflect not only conductive but also capacitive properties and allow to derive the epithelial capacitance, which directly depends on the epithelial surface area [12]. Impedance spectroscopy uses the fact that the plasma membrane of epithelial cells acts as a capacitor (C^{epi}) that causes a phase shift between current and voltage under alternate current (AC) conditions. C^{epi} is short-circuited by R^{epi}, resulting in a time constant $\tau = R^{epi} \cdot C^{epi}$. With R^{epi} of different cell types and tissues varying between about 10 and 10.000 Ωcm^2 and C^{epi} values of about 1 to 10 $\mu F/cm^2$, τ ranges between 10^{-5} and 10^{-1} s. In terms of angular frequency ($\omega = 1/\tau$) this is equal to a range from 10 Hz to 100 kHz. During a typical experiment, up to 50 different frequencies covering this range are used to obtain impedance values $|Z|$ and corresponding phase angles ϕ [15].

Impedance spectroscopy was already used during the first half of the 20th century [16][17], and gained popularity in epithelial research since the 1970s (see e.g., [18][19][20][21]). To analyze samples, an equivalent electric circuit of appropriate complexity is modeled [22]. The simplest circuit that incorporates R^{epi} is a resistor-capacitor (RC) circuit (Figure 2b). In a previous publication, the evaluation of impedance spectra based on this equivalent circuit was dubbed one-path impedance spectroscopy (1PI) to distinguish it from the more recently developed two-path impedance spectroscopy (2PI, [23][13]) that additionally allows the separation of the two major transepithelial transport pathways (trans- and paracellular). In 2PI, a circuit with an RC subcircuit and a resistor in parallel is used to represent the physiological polarity of

epithelial cells (Figure 2c). In both cases, the subepithelium can be represented by an additional resistor in series.

In theory, R^{sub} and R^{epi} can be derived by physical relationships. In particular, $\lim_{\omega\to\infty} \Re(Z) = R^{sub}$ and $\lim_{\omega\to 0} \Re(Z) = R^{sub} + R^{epi}$ can be exploited for this task. In previous work, we have demonstrated that traditional ways to estimate R^{epi} from these relations, like circle fits or visual extrapolation from Nyquist diagrams [22], can lead to serious errors in the analysis of epithelial characteristics [13]. This holds true, in particular, where spectra deviate from a semicircular shape. We have also demonstrated that estimations can be improved by applying machine learning techniques on complex-valued impedances of error-prone [13] or on extracted features of ideal impedance spectra, respectively [24]. For R^{epi}, e.g., we achieved less than ten percent deviation from the known target value on average with outliers, however, exhibiting significantly larger maximum deviations [24]. Recently, we have tested a novel approach to quantify electric properties of epithelia based on extracted features of error-prone impedance spectra [25].

C^{epi} is established by the transepithelial pathway or the hydrophobic part of the lipid bilayers of its cell membranes, respectively. Under the AC conditions induced by impedance measurements, charging and discharging of the epithelial capacitor depends on frequency. At very high frequencies the impedance of the capacitor approaches zero [26]. In previous work, we have shown that this physical relation can be used to approximate C^{epi} from the cartesian representation of an impedance spectrum [12]. The main advantage of this is the minimal requirements for additional equipment. Alternative methods employ square pulses and the analysis of resulting output transients [17][27] or very brief pulses or simultaneously imposing different sine waves and using Fourier analysis [28][29]. Currently, however, estimations of C^{epi} by machine learning are less precise than those by more complex and more time-consuming methods [12].

The term machine learning generally comprises a variety of algorithms that are characterized by learning a function from given samples. Depending on the employment or non-employment of target values for each sample, these algorithms are either regarded as supervised or unsupervised learning algorithms [30]. To this end, quantification of epithelial properties is a supervised learning task. Among the most widely used supervised machine learning techniques are neural net-
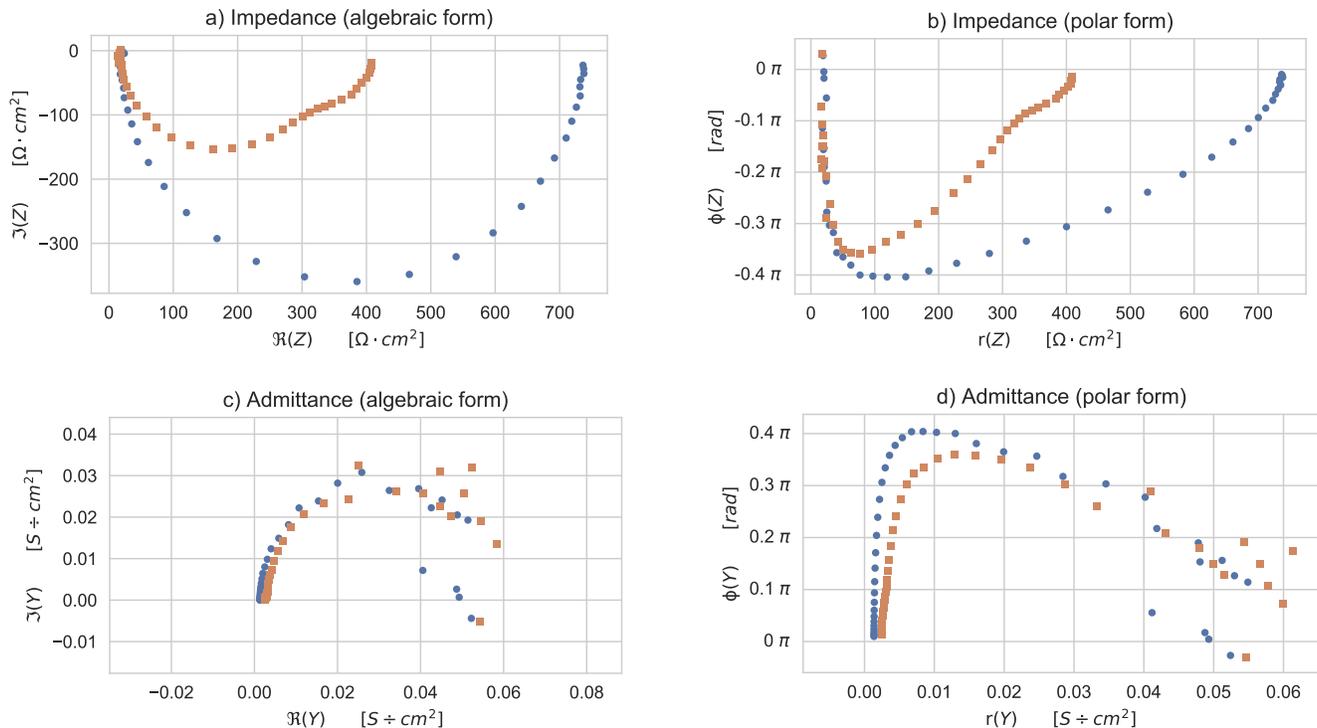
Figure 3. Overlay of two impedance measurements of modeled HT29B6 at 42 frequencies using different representations. Colors indicate functional condition before (●) and after (■) Nystatin addition. a) Nyquist representation: impedance in algebraic form ($\Re(Z)$, $\Im(Z)$) b) impedance in polar form ($r(Z)$, $\phi(Y)$) c) admittance in algebraic form ($\Re(Y)$, $\Im(Y)$) (Nyquist representation) d) admittance in polar form ($r(Y)$, $\phi(Y)$).

works, which are biologically motivated, and random forests, which are statistically motivated. Neural networks model the characteristics of biological neurons as mathematical functions [31] and have been proven to be universal approximators [32]. Random Forests are based on the concept of decision trees [33] and exploit the mathematical effects of induced variability among a large number of decision trees [34].

As a routine preprocessing task, many machine learning applications employ an extraction and/or selection of input features. While feature extraction typically aims at transforming or increasing existing features [35], feature selection can be considered as an algorithmic dimension reduction process. Feature selection techniques are typically either purely statistically motivated filter approaches or so-called wrapper approaches that make use of supervised learning algorithms [36]. Well-known wrapper approaches are, e.g., Sequential Forward Selection or Recursive Feature Elimination [37]. Apart from these general strategies, embedded feature selection approaches are known as effective strategies as they are implemented directly into supervised learning algorithms [36]. Random Forests are prominent example for embedded feature selection [34].

To achieve reliable predictions, machine learning requires large amounts of training data. As measurements are typically only available in small numbers ($n \leq 1.000$), employing this approach to epithelial analysis requires modeling realistic impedance data. Therefore, we have established a modeling methodology, which is based on an equivalent electric circuit reflecting epithelial polarity (Figure 2d). This includes not only parameter ranges for circuit components and a mathematical model of setup-specific data scatter, but also an algorithmic

approach to compare modeled and measured data [13][25]. Detailed and realistic models have been developed for three epithelial cell lines under various conditions [1]: the human colon carcinoma cell line HT-29/B6, the porcine jejunum cell line IPEC-J2, and the canine kidney cell line MDCK I.

Figure 3a shows an example of the impedance spectrum of a modeled HT29B6 cell culture under control conditions and after apical addition of Nystatin. Note, how the shapes of the two curves differ: Whereas under control conditions an approximately semicircular curve is seen, the shape after addition of Nystatin can be described as two overlapping semicircles representing the two cell membrane sides. This is due to the reduction of the apical membrane resistance R^{ap} which leads to a difference in the time constants of the apical membrane τ^{ap} and basolateral membrane τ^{bl}. As an indicator of the curve shape, previous work has predicted the τ ratio q, where impedance spectra can be considered semicircular for $q < 5$ and non-semicircular for $q > 5$ [38].

In addition to the use of Cartesian coordinates, i.e., the algebraic form, other forms of representation can be considered for the analysis of impedance measurements. Therefore, polar coordinates of impedance have been calculated in previous studies to predict epithelial properties using machine learning (e.g., in [39][12][1]). In addition to impedance, its complex inverse, the admittance, can also be considered, which can also be represented in coordinates and polar form. In the case of admittance, in particular, there is a lack of more detailed research on how well it can be used for machine-learning-based analyses. Figure 3 shows the four representations and the influence of an apical addition of Nystatin.

III. Modeling Impedance Spectra

For the present study, an extensive dataset was synthesized that mimics impedance measurements on the three cell lines HT-29/B6, IPEC-J2, and MDCK I. The data is freely available online at [40]. For each sample, two different functional states are considered together: control conditions and application of Nystatin. By the modeling procedures described in the following, we produced 150,000 samples per cell line and tissue condition (cf. Table I). As described in the following, the modeling process comprised assumptions about an underlying equivalent electric circuit for both functional states as well as calculating impedances for given frequencies and modeling equipment-specific measurements biases.

TABLE I. SAMPLE SIZES FOR MODELED EPITHELIAL CELL LINES AND THEIR FUNCTIONAL CONDITIONS.

Condition	HT-29/B6	IPEC-J2	MDCK I
Control	150.000	150.000	150.000
Nystatin	150.000	150.000	150.000

A. Modeling Impedance Spectra

For all three modeled cell lines, an equivalent circuit consisting of two RC subcircuits a (R_a, C_a) and b (R_b, C_b) located in series, a resistor in parallel (R_p) and a concluding resistor in series (R_s) is assumed.(cf. Figure 2). In accordance with Kirchhoff's laws, the corresponding impedance Z at an angular frequency ω can be derived from the complex-valued impedance of the circuit components:

$$Z(\omega) = \frac{R_p(R_a + R_b) + i\omega[R_p(R_a\tau_b + R_b\tau_a)]}{R_a + R_b + R_p(1 - \omega^2\tau_a\tau_b) + i\omega[R_p(\tau_a + \tau_b) + R_a\tau_b + R_b\tau_a]} + R_s \quad (2)$$

where $i = \sqrt{-1}$, and $\tau_a = R_aC_a$ and $\tau_b = R_bC_b$.

In the measurements to be modeled, ten frequencies per decade are used. Based on a lowest frequency f_1 of 1.3 Hz, higher frequencies f_i with $1 < i \leq n = 42$ are multiples by a factor of $10^{0.1}$ (~1.26). Note, that the value of f_1 is a chosen to a avoid obtaining multiples of 50 Hz (mains frequency) and that for application with Eq. (2), $\omega_i = 2\pi/f_i$ is calculated.

From calculations, $n=42$ complex-valued impedances $(Z_0, \ldots, Z_{n-1})$ are obtained. Real and imaginary parts of a spectrum can be regarded as separate feature sets $S_{\Re(Z)}$ and $S_{\Im(Z)}$:

$$S_{\Re(Z)} = \{\Re(Z_0), \ldots, \Re(Z_{n-1})\} \quad (3)$$
$$S_{\Im(Z)} = \{\Im(Z_0), \ldots, \Im(Z_{n-1})\} \quad (4)$$

As an alternative representation, these complex values were transformed to polar coordinates, i.e., to phase angle ϕ and magnitude r. This results in two alternative sets S_ϕ and S_r:

$$S_{\phi(Z)} = \{\phi(Z_0), \ldots, \phi(Z_{n-1})\} \quad (5)$$
$$S_{r(Z)} = \{r(Z_0), \ldots, r(Z_{n-1})\} \quad (6)$$

Furthermore, the admittances $Y_i = Z_i^{-1}$, i.e., the complex reciprocals of the impedances, were computed for the whole spectrum. Feature sets $S_{\Re(Y)}, S_{\Im(Y)}$ correspond to their real and imaginary parts.

$$S_{\Re(Y)} = \{\Re(Y_0), \ldots, \Re(Y_{n-1})\} \quad (7)$$
$$S_{\Im(Y)} = \{\Im(Y_0), \ldots, \Im(Y_{n-1})\} \quad (8)$$

As for the impedance, the polar form of the admittance was derived from the algebraic coordinates which leads to feature sets $S_{\phi(Y)}, S_{r(Y)}$

$$S_{\phi(Y)} = \{\phi(Y_0), \ldots, \phi(Y_{n-1})\} \quad (9)$$
$$S_{r(Y)} = \{r(Y_0), \ldots, r(Y_{n-1})\} \quad (10)$$

Thus, we obtain four different representations of a spectral impedance measurement, which can be seen in Figure 3.

B. Modeling Data Scatter

Further, we model and apply synthetic data scatter reflecting deviations from the theoretical impedance value caused by the electrophysiological measurement setup. This scatter is modeled based on relative deviations of the real part $\Re$ and imaginary part $\Im$ of measured impedances from theoretical values. For a given impedance Z at frequency $f = \omega/2\pi$, the relative deviation σ of $\Re(Z)$ is approximated as second-order Fourier series (n=2) and relative deviation of $\Im(Z)$ is approximated as fourth-order polynomial function (n=4):

$$\sigma_\Re(f) = a_0 + \sum_{i=1}^{n} a_i \cdot cos(nwf) + b_i \cdot sin(nwf) \quad (11)$$

$$\sigma_\Im(f) = a_0 + \sum_{i=1}^{n} a_i \cdot f^i \quad (12)$$

where coefficients w, a_0, a_1, b_1, a_2, b_2 or a_0, a_1, a_2, a_3 and a_4 were determined by function fitting. a_0 is modeled as dependent on R^T. For further details on the data scatter model, see [13] or [41].

For all synthetic samples used in the following, complex impedances are calculated according to the model parameters. Subsequently, data scatter is added, and finally, polar impedances and admittances are calculated from the scattered complex impedances.

Note, that for the imaginary parts of the impedance, the modeled scatter can lead to positive values at measurement points that were already close to zero (cf. Figure 3a). This in turn leads to negative imaginary parts of the admittance (cf. Figure 3c). This modeling agrees with our observations from laboratory practice, although in theory such values should not occur (cf. Eq. (2)).

C. Sampling Cell Lines for Control Conditions

While IPEC-J2 and MDCK I cells typically show relatively high R^{epi} values under physiological conditions, for HT-29/B6 cells relatively low values are measured. Based on this, as well as on further published measurement results, parameter ranges of the components of the equivalent circuit had been defined in previous work [25]. An overview of the parameter ranges for the synthesis is shown in Table II. In an early synthesis phase, samples with biologically implausible values were obtained. Under the control conditions it is therefore presumed that $R^{ap} > R^{bl}$ applies to all cell lines and $\frac{1}{4} < \frac{R^{para}}{R^{trans}} < 4$ applies to HT-29/B6 and MDCK I.

From the observation that impedance spectra from epithelia usually show a symmetric semi-circular shape under control conditions, it can be concluded that apical and basolateral polarity outbalance each other [22]. In other words, the τ ratio q is expected to be close to 1. Therefore, impedance spectra, whose equivalent circuits have a τ-ratio greater than five, were

TABLE II. MODELED PARAMETER RANGES FOR CELL LINES
UNDER CONTROL CONDITIONS

Parameter	HT-29/B6		IPEC-J2		MDCK I		unit
	min	*max*	*min*	*max*	*min*	*max*	
R^{sub}	8	30	8	50	8	25	$\Omega \cdot cm^2$
R^{epi}	150	1498	900	8567	100	4495	$\Omega \cdot cm^2$
R^{para}	152	30000	944	15000	102	10000	$\Omega \cdot cm^2$
R^{trans}	151	20000	3000	20000	101	15000	$\Omega \cdot cm^2$
R^{ap}	1	19500	2500	19000	10	14500	$\Omega \cdot cm^2$
C^{epi}	1.0	5.5	0.7	2.0	0.8	3.3	$\mu F/cm^2$
C^{ap}	1.3	8.1	1.1	3.3	0.9	5.7	$\mu F/cm^2$
C^{bl}	5.0	86.2	1.5	9.3	2.9	9.5	$\mu F/cm^2$

filtered out. Visual inspection confirmed that this yields only impedance spectra with an almost semicircular shape.

The obtained dataset shows different distributions of the equivalent circuit parameters for the three cell lines, as can be seen in Figure 4. For example, it is clear that for HT-29/B6 modeling, as in reality, significantly smaller values for R^{epi} occur than for MDCK I and IPEC-J2 (cf. Figure 4a). Note, that for HT-29/B6 higher capacitances occur, especially on the basolateral side (cf. Figure 4d). MDCK I have on average slightly smaller resistances than IPEC-J2 (cf. Figure 4a-c), but slightly larger capacitances (cf. Figure 4d).

D. Modeling Nystatin

For estimating apical and basolateral cell membrane properties, a one-sided addition of the substance Nystatin has been used, as it increases the conductance on one side of the tissue without noticeably changing the conductance on the other side [42]. For the equivalent circuit, the apical application of Nystatin means a decrease of the apical resistance R^{ap}. To model the application of the drug to our cells, the Nystatin factor $d_{nyst} \in [0; 1]$ is introduced. The apical resistance after Nystatin application R_2^{ap} is calculated according to the following formula:

$$R_2^{ap} = R^{ap} \cdot d_{nyst} \quad (13)$$

Note, that the stronger the effect of Nystatin, the lower d_{nyst}. In reality, the effect of Nystatin depends on the amount of substance added and the time between application and measurement. The reduction of the apical resistance also results in a reduced transcellular resistance R_2^{trans}, epithelial resistance R_2^{epi} and apical time constant τ_2^{ap}.

In laboratory practice, Nystatin is applied until two distinguishable semicircles are visible in the Nyquist representation. In this study this procedure is modeled as a multi-step Nystatin application. For a series of k Nystatin applications, the Nystatin factor is calculated from the product of the individual applications.

$$d_{nyst} = \prod_{i=1}^{k} d_{nyst}^i \quad (14)$$

Note, that $d_{nyst}^i = 1.0$ means no change and a value of 0 means complete removal of apical resistance. Therefore, at each step, a random, uniformly distributed value is chosen from the interval $[0.2; 0.8]$ before updating the apical resistance R_2^{ap}. This is repeated until a τ ratio above five is reached, ensuring two distinguishable semicircles in the Nyquist representation.
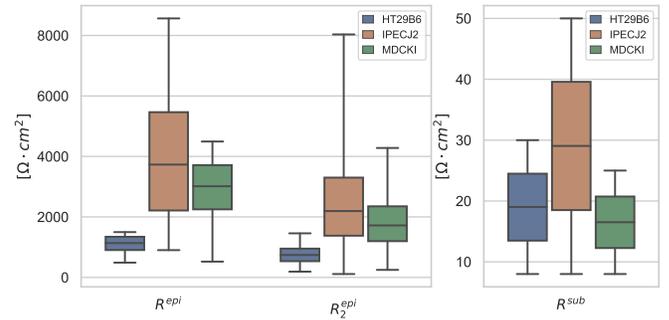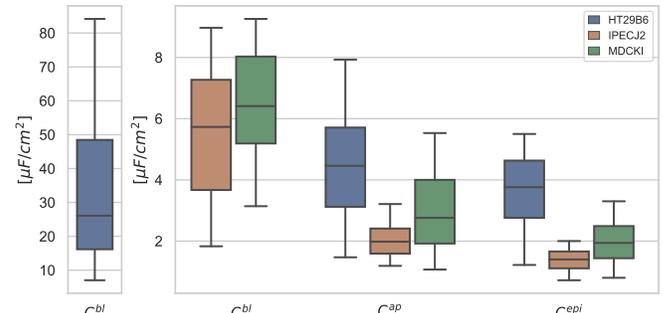

a) Epithelial resistance before (R^{epi}) and after Nystatin application (R_2^{epi}) and subepithelial resistance (R^{sub}).
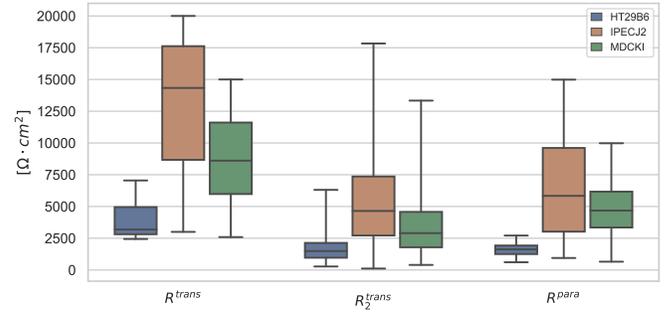
b) Transcellular resistance before (R^{trans}) and after Nystatin application (R_2^{trans}) and paracellular resistance (R^{para}).
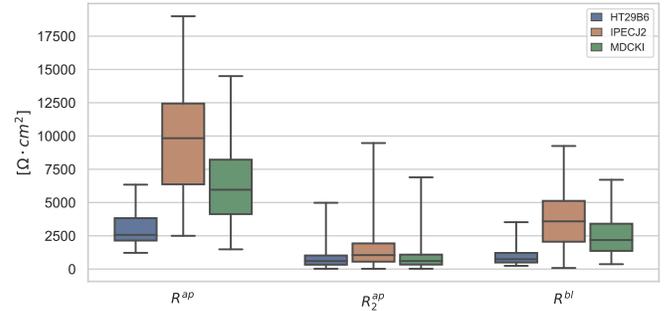
c) Apical resistance before (R^{ap}) and after Nystatin application (R_2^{ap}) and basolateral resistance (R^{bl}).

d) Apical capacity (C^{ap}), basolateral capacity (C^{bl}) and epithelial capacity (C^{epi}).

Figure 4. Overview of the distribution of the target parameters for the modeled cell lines as box plots with whiskers showing the minimum and maximum values.

TABLE III. INPUT FEATURE SETS COMPRISING IMPEDANCE SPECTRA AND STATISTICAL FEATURES.

Feature Set	Definition	Approach	Size	Quality	Form	Functional States	Ratios and Differences*
$I^1_{\perp(Z)}$	$S^*_{\Re(Z)} \times S^*_{\Im(Z)}$	1	106	Impedance	algebraic	Control	–
$I^1_{\angle(Z)}$	$S^*_{\phi(Z)} \times S^*_{r(Z)}$	1	106	Impedance	polar	Control	–
$I^1_{\perp(Y)}$	$S^*_{\Re(Y)} \times S^*_{\Im(Y)}$	1	106	Admittance	algebraic	Control	–
$I^1_{\angle(Y)}$	$S^*_{\phi(Y)} \times S^*_{r(Y)}$	1	106	Admittance	polar	Control	–
$I^2_{\perp(Z)}$	$I^1_{\perp(Z)} \times N^*_{\Re(Z)} \times N^*_{\Im(Z)}$	2	212	Impedance	algebraic	Control + Nystatin Affected	–
$I^2_{\angle(Z)}$	$I^1_{\angle(Z)} \times N^*_{\phi(Z)} \times N^*_{r(Z)}$	2	212	Impedance	polar	Control + Nystatin Affected	–
$I^2_{\perp(Y)}$	$I^1_{\perp(Y)} \times N^*_{\Re(Y)} \times N^*_{\Im(Y)}$	2	212	Admittance	algebraic	Control + Nystatin Affected	–
$I^2_{\angle(Y)}$	$I^1_{\angle(Y)} \times N^*_{\phi(Y)} \times N^*_{r(Y)}$	2	212	Admittance	polar	Control + Nystatin Affected	–
$I^3_{\perp(Z)}$	$I^2_{\perp(Z)} \times D^*_{\Re(Z)} \times D^*_{\Im(Z)} \times R^*_{\Re(Z)} \times R^*_{\Im(Z)}$	3	318	Impedance	algebraic	Control + Nystatin Affected	included
$I^3_{\angle(Z)}$	$I^2_{\angle(Z)} \times D^*_{\phi(Z)} \times D^*_{r(Z)} \times R^*_{\phi(Z)} \times R'^*_{r(Z)}$	3	318	Impedance	polar	Control + Nystatin Affected	included
$I^3_{\perp(Y)}$	$I^2_{\perp(Y)} \times D^*_{\Re(Y)} \times D^*_{\Im(Y)} \times R^*_{\Re(Y)} \times R'^*_{\Im(Y)}$	3	318	Admittance	algebraic	Control + Nystatin Affected	included
$I^3_{\angle(Y)}$	$I^2_{\angle(Y)} \times D^*_{\phi(Y)} \times D^*_{r(Y)} \times R^*_{\phi(Y)} \times R'^*_{r(Y)}$	3	318	Admittance	polar	Control + Nystatin Affected	included

*as calculated in section III-E

Figure 4 shows the distributions of equivalent circuit parameters for all cell lines after modeling Nystatin in this manner.

After updating the circuit parameters, new impedance spectra were calculated based on the modified equivalent circuit. Thus, for each sample, we obtain two impedance spectra at two different functional states. As for the state under control conditions, the feature sets for the state after Nystatin application were computed according to the formulas 3-10, but are denoted as N_κ in the following.

E. Extracted Features

New feature sets can be constructed by considering the impedance spectra before and after Nystatin application. In order to describe the change of spectra due to the application of Nystatin, differences and ratios of the measuring points were calculated for all representations, resulting in the feature sets D_κ and R_κ.

$$D_\kappa = \{n - s \mid n \in N_\kappa \wedge s \in S_\kappa \wedge \omega(n) = \omega(s)\} \quad (15)$$
$$R_\kappa = \{n / s \mid n \in N_\kappa \wedge s \in S_\kappa \wedge \omega(n) = \omega(s)\} \quad (16)$$

where $\kappa \in \{\Re(Z), \Im(Z), \phi(Z), r(Z), \Re(Y), \Im(Y), \phi(Y), r(Y)\}$ and $\omega(x)$ is the angular frequency for the measurement of x.

Previous work has shown that the use of additional statistical features can lead to an improved determination of epithelial properties [1]. According to this principle, 11 statistical properties were selected (cf. Table IV). The function stat : $\mathbb{R}^{42} \to \mathbb{R}^{11}$ returns these statistical features for a given feature set. For all previously defined feature sets S_κ, N_κ, D_κ, and R_κ, the inclusion of statistical features results in new feature sets S^*_κ, N^*_κ, D^*_κ, and R^*_κ.

$$S^*_\kappa = S_\kappa \times \text{stat}(S_\kappa) \quad (17)$$
$$N^*_\kappa = N_\kappa \times \text{stat}(N_\kappa) \quad (18)$$
$$D^*_\kappa = D_\kappa \times \text{stat}(D_\kappa) \quad (19)$$
$$R^*_\kappa = R_\kappa \times \text{stat}(R_\kappa) \quad (20)$$

where $\kappa \in \{\Re(Z), \Im(Z), \phi(Z), r(Z), \Re(Y), \Im(Y), \phi(Y), r(Y)\}$

TABLE IV. STATISTICAL PROPERTIES FOR FEATURE SET S_i.

Feature	Definition with $x_j \in S_i$ and $n = \#(S_i)$	Description
$min(S_i)$	$\{x : x \leq x_j \ \forall \ x_j \in S_i\}$	Minimum
$P_{10}(S_i)$	$x_{\lfloor 0.1 \cdot n + 1 \rfloor}$	10th Percentile
$P_{25}(S_i)$	$x_{\lfloor 0.25 \cdot n + 1 \rfloor}$	1st Quartile
$med(S_i)$	$\begin{cases} x_{(\frac{n+1}{2})}, & \text{if } n \text{ odd} \\ \frac{1}{2}(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}), & \text{if } n \text{ even} \end{cases}$	2nd Quartile (Median)
$P_{75}(S_i)$	$x_{\lfloor 0.75 \cdot n + 1 \rfloor}$	3th Quartile
$P_{90}(S_i)$	$x_{\lfloor 0.9 \cdot n + 1 \rfloor}$	90th Percentile
$max(S_i)$	$\{x : x \geq x_j \ \forall \ x_j \in S_i\}$	Maximum
$\bar{x}(S_i)$	$\frac{1}{n} \sum_{j=1}^{n} x_j$	Arithmetic mean
$s(S_i)$	$\sqrt{s^2(S_i)}$	Standard deviation
$s^2(S_i)$	$\frac{1}{n-1} \sum_{j=1}^{n} (x_j - \bar{x}(S_i))^2$	Variance
$R_{MM}(S_i)$	$med(S_i) - \bar{x}(S_i)$	Distance between median and arithmetic mean

F. Comparison with Measured Data

As outlined (cf. Table I), modeling the cell lines HT-29/B6, IPEC-J2 and MDCK I under two distinct conditions, six modeling scenarios were considered. For all six scenarios, modeled impedance spectra were compared to measured spectra. As true values can not be known for measured parameters, an indirect approach needs to be chosen that involves differences between estimations. Such an indirect approach has been proposed [13] and optimized [1] by us in previous work: Each parameter is estimated by three alternative techniques, normalized and resulting estimation differences were plotted in a two-dimensional diagram for modeled and measured data, respectively. If data points from measured spectra lie within the area covered by the data obtained from model spectra, it is assumed that the modeling reflects realistic epithelial properties. By calculating hexagonal bins for both diagrams, the overlap or similarity was quantified. For further details on the calculation and quantification of similarity, see [41].

IV. Quantifying Electric Properties by Machine Learning

This section describes how physiological properties of epithelia can be determined by applying machine learning techniques on impedance measurements. In particular, we investigate how much the predictions of the equivalent circuit parameters can be improved by a two-state consideration. First, necessary preprocessing and feature selection steps are described. In the main step, parameters of the tissue-equivalent electric circuit model are determined by separate regression models (cf. Figure 1). Therein, the known parameter values from the modeled impedance data set are used for supervised learning and subsequent evaluation.

A. Preprocessing and Feature Selection

As described in the previous section, varying input feature sets were formed from the cell measurements under control conditions and their statistical properties. In particular, different qualities and forms of measurements were considered and combined. In general, this approach is analogous to our previous work [1].

In addition to features from control conditions data, the pair of spectra, i.e., before and after Nystatin application, and statistical properties were used to extract features. The last approach results from combining the features from the second approach and the differences and ratios of the measurement points under the two states as described before. For each of the three approaches, the four representations of impedance measurements were used.

All in all, this results in 12 input feature sets, which can be seen in Table III. To determine which features are suitable for estimating the physiological properties of epithelia, the feature sets were evaluated by assessing the results of various supervised learning procedures.

To evaluate the feature sets and learning methods, the absolute percentage error *APE* is calculated for all test samples from the target value t and its prediction $\hat{t}$:

$$APE(t, \hat{t}) = \frac{|\hat{t} - t|}{t} \tag{21}$$

Samples from each cell line were randomized and standardized by removing the mean and scaling to unit variance. Subsequently, the dataset was divided to obtain a training set of 100,000 and a test set of 50,000 samples.

B. Supervised Learning

While the general approach aims to determine several circuit parameters, individual circuit parameters were determined instead of multiple regression. With twelve feature sets, six targets, and three machine learning models, the total number of trained models is 216 per cell line. For the whole study, 648 models were trained.

This approach allowed to choose the best model and input feature set for each parameter. Thus, the values of R^{sub}, R^{para}, R^{ap}, R^{bl}, C^{ap}, C^{bl} under control conditions were determined. To quantify the effect of Nystatin addition, d_{nyst} was also determined, which was then used to calculate R_2^{ap} using the prediction of R^{ap}.

Previous work has shown that machine learning methods, such as random forests and multilayer perceptrons, provide more accurate estimates than traditional methods such as circle fit [12][1]. On this basis, three machine learning regression methods of varying complexity were used for this work:

- **Decision Tree (DT).** This simple regressor serves as a baseline regression method. The depth of the tree is not limited and two samples are sufficient to split an inner node. The best split is always chosen based on the mean squared error.

- **Random Forest (RF.)** The statistically motivated regressor was employed using an ensemble of 500 decision trees. As with single decision tree regression, the random forest trees are not limited in-depth and a split requires at least two samples. The best split is chosen based on the mean squared error.

- **Multilayer Perceptron (MLP).** This biologically inspired neural network regressor was used with the seven-layers network architecture $[m, 256, 128, 64, 32, 16, 1]$, where m denotes the dimensionality of the input vector. Thus, the number of input neurons depends on the input feature set. Rectified linear units (ReLU) are used as the activation function. For training, the Adam optimizer of Kingma et al. was used [43]. This is a variant of the stochastic gradient descent method with backpropagation and adaptive learning rate. The initial learning rate was 0.001. In addition, an L2 regularization term with an α of 0.0001 was used. 200 samples were combined to a mini-batch.

Focussing on selecting effective feature sets, extensive optimization of hyperparameters was omitted for the regression methods at this point, although these were chosen based on experience from past work. For all supervised learning tasks, machine learning methods were used as provided by the Python library *scikit-learn* [44].

C. Results

By comparing prediction results from the 648 trained machine learning models, we identified feature sets best suited for the given task of determining R^{ap}, R^{bl}, etc (Table V). The underlying assumption was that prediction accuracies will differ depending on the input feature set and the machine learning model used.

From all 648 trained models, we show results for the best 162 models in Table V. For all target variables, regression models were considered that estimate with a MAPE of less than 5%. For every single approach and target, the feature set that leads to the minimum *MAPE* value was selected. The arithmetic mean over the APE values of all test samples *MAPE* is a relative metric independent of actual target values and ranges. It is particularly well suited as a measure to compare the prediction accuracies among different target variables.

For all cell lines, targets, and machine learning models, the feature sets of the second approach perform better than the first approach (cf. Table V). This is especially evident in the estimation of apical and basolateral features. In many cases, the *MAPE* can further be reduced by adding extra features using the third approach. For example, in predicting C^{bl} for IPEC-J2, the first approach achieves a *MAPE* of 18.7%, while the *MAPE* for the second approach is 2.5% and 2.0% for the third approach (cf. Table V).

TABLE V. MAPE FOR PREDICTING EPITHELIAL PROPERTIES [±%].

a) Estimation of C^{ap}

Cell Line	Approach	Feature Set	DT	RF	MLP
HT29B6	1	$I^1_{\perp(Y)}$	8.43	7.01	6.90
	2	$I^2_{\angle(Z)}$	2.88	1.84	**0.66**
	3	$I^3_{\perp(Y)}$	**2.60**	**1.64**	0.67
IPECJ2	1	$I^1_{\perp(Y)}$	9.77	8.13	7.73
	2	$I^2_{\perp(Y)}$	**3.82**	**2.79**	1.02
	3	$I^3_{\perp(Y)}$	2.85	1.92	1.31
MDCKI	1	$I^1_{\perp(Y)}$	5.58	4.79	5.84
	2	$I^2_{\perp(Y)}$	2.87	2.10	0.93
	3	$I^3_{\perp(Y)}$	**2.24**	**1.50**	**0.77**

c) Estimation of R^{sub}

Cell Line	Approach	Feature Set	DT	RF	MLP
HT29B6	1	$I^1_{\perp(Z)}$	3.24	2.22	1.42
	2	$I^2_{\angle(Z)}$	**2.65**	**1.76**	1.31
	3	$I^3_{\angle(Z)}$	2.67	1.78	**1.27**
IPECJ2	1	$I^1_{\perp(Z)}$	3.75	2.56	2.05
	2	$I^2_{\perp(Z)}$	**3.02**	**1.96**	**1.32**
	3	$I^3_{\perp(Z)}$	3.06	1.98	2.66
MDCKI	1	$I^1_{\perp(Z)}$	4.88	3.36	2.49
	2	$I^2_{\perp(Z)}$	**3.90**	**2.60**	**2.45**
	3	$I^3_{\perp(Z)}$	3.95	2.62	2.49

e) Estimation of R^{bl}

Cell Line	Approach	Feature Set	DT	RF	MLP
HT29B6	1	$I^1_{\perp(Y)}$	50.96	44.06	34.81
	2	$I^2_{\angle(Z)}$	5.54	5.26	2.34
	3	$I^3_{\perp(Z)}$	**3.06**	**2.98**	**1.99**
IPECJ2	1	$I^1_{\perp(Y)}$	44.87	36.85	25.07
	2	$I^2_{\perp(Z)}$	10.35	8.19	4.90
	3	$I^3_{\perp(Y)}$	**6.04**	**5.26**	**3.56**
MDCKI	1	$I^1_{\perp(Y)}$	35.08	27.51	22.66
	2	$I^2_{\perp(Z)}$	6.66	5.59	1.93
	3	$I^3_{\angle(Y)}$	**4.42**	**3.78**	**1.78**

b) Estimation of C^{bl}

Cell Line	Approach	Feature Set	DT	RF	MLP
HT29B6	1	$I^1_{\perp(Y)}$	45.13	40.73	33.33
	2	$I^2_{\perp(Z)}$	9.25	5.56	**1.65**
	3	$I^3_{\perp(Z)}$	**8.76**	**5.44**	2.02
IPECJ2	1	$I^1_{\perp(Y)}$	24.53	20.49	18.70
	2	$I^2_{\perp(Z)}$	8.37	5.62	2.53
	3	$I^3_{\perp(Y)}$	**5.71**	**3.72**	**2.00**
MDCKI	1	$I^1_{\perp(Y)}$	12.86	11.09	11.53
	2	$I^2_{\perp(Z)}$	6.66	4.38	1.85
	3	$I^3_{\perp(Y)}$	**5.55**	**3.49**	**1.09**

d) Estimation of R^{ap}

Cell Line	Approach	Feature Set	DT	RF	MLP
HT29B6	1	$I^1_{\perp(Y)}$	20.73	17.03	13.43
	2	$I^2_{\perp(Z)}$	6.03	6.43	2.58
	3	$I^3_{\perp(Z)}$	**4.46**	**4.76**	**2.21**
IPECJ2	1	$I^1_{\perp(Y)}$	37.64	31.55	22.92
	2	$I^2_{\perp(Y)}$	22.60	18.73	4.85
	3	$I^3_{\perp(Y)}$	**16.45**	**15.11**	**4.25**
MDCKI	1	$I^1_{\perp(Y)}$	31.11	24.39	19.74
	2	$I^2_{\perp(Z)}$	10.48	10.28	**2.93**
	3	$I^3_{\perp(Z)}$	**8.78**	**8.59**	3.04

f) Estimation of R^{para}

Cell Line	Approach	Feature Set	DT	RF	MLP
HT29B6	1	$I^1_{\perp(Z)}$	7.25	6.95	5.43
	2	$I^2_{\perp(Z)}$	2.57	2.59	**0.75**
	3	$I^3_{\perp(Z)}$	**2.07**	**2.03**	1.07
IPECJ2	1	$I^1_{\perp(Y)}$	17.36	13.82	9.76
	2	$I^2_{\perp(Z)}$	7.81	6.84	**3.38**
	3	$I^3_{\perp(Z)}$	**6.94**	**5.90**	4.26
MDCKI	1	$I^1_{\perp(Y)}$	16.08	13.02	9.57
	2	$I^2_{\perp(Z)}$	5.67	4.93	2.16
	3	$I^3_{\perp(Z)}$	**5.00**	**4.17**	**1.36**

TABLE VI. SELECTED ROBUST FEATURE SETS

Target	HT-29/B6 Feature Set	Model	IPEC-J2 Feature Set	Model	MDCK I Feature Set	Model
C^{ap}	$I^2_{\perp(Y)}$	MLP	$I^2_{\angle(Z)}$	MLP	$I^2_{\angle(Z)}$	MLP
C^{bl}	$I^2_{\angle(Y)}$	MLP	$I^2_{\angle(Z)}$	MLP	$I^2_{\angle(Z)}$	MLP
R^{sub}	$I^2_{\perp(Z)}$	MLP	$I^2_{\perp(Z)}$	MLP	$I^2_{\perp(Z)}$	MLP
R^{ap}	$I^2_{\angle(Z)}$	RF	$I^1_{\perp(Y)}$	MLP	$I^3_{\angle(Y)}$	RF
R^{bl}	$I^2_{\angle(Y)}$	MLP	$I^1_{\perp(Y)}$	MLP	$I^3_{\angle(Y)}$	RF
R^{para}	$I^2_{\perp(Y)}$	RF	$I^2_{\angle(Y)}$	MLP	$I^3_{\perp(Y)}$	RF
R^{ap}_2	$I^2_{\angle(Z)}$	MLP	$I^3_{\perp(Y)}$	RF	$I^2_{\angle(Y)}$	MLP

TABLE VII. ABSOLUTE PERCENTAGE ERROR FOR MACHINE LEARNING WITH ROBUST FEATURE SETS

	HT-29/B6				IPEC-J2				MDCK I			
	mean	med	$P_{95\%}$	max	mean	med	$P_{95\%}$	max	mean	med	$P_{95\%}$	max
C^{ap}	0.7	0.5	1.8	9.1	1.3	0.9	3.7	20.0	0.8	0.7	2.2	12.0
C^{bl}	1.2	0.9	3.4	35.9	2.1	1.6	5.9	34.2	1.3	1.0	3.4	31.9
R^{sub}	1.3	1.0	3.7	9.9	1.3	0.9	4.1	20.0	2.3	1.7	6.4	20.7
R^{ap}	8.8	4.8	30.5	91.6	4.8	2.5	15.5	303.0	8.6	3.4	36.6	176.9
R^{bl}	1.5	1.0	4.1	38.2	3.3	2.1	10.5	99.8	3.8	1.8	14.6	56.5
R^{para}	4.7	3.7	13.6	29.2	3.3	1.9	10.6	115.4	6.9	4.3	23.1	67.5
R^{ap}_2	10.2	6.4	32.7	152.8	14.9	7.9	52.4	320.4	11.8	7.8	36.4	215.2

Among the machine learning models, the MLPs lead to the smallest *MAPE* values. In comparison to the baseline method of the decision tree, however, the random forest also yielded an improvement of the estimates in the vast majority of cases.

D. Outlier and Robust Feature Sets

Even though it could be shown that machine learning may yield a very accurate determination of epithelial properties, in a few cases very high relative deviations occur. In practice, however, even a few extreme outliers may be more harmful than lower average accuracies. To this end, an approach that avoids extreme outliers is favorable and a reliable method is needed that provides useful estimates even in the worst cases.

Therefore, in an alternative approach, feature sets and machine learning models were selected for which the smallest relative deviations (i.e., APE values) occur in the worst case within the test samples. These alternative feature sets for all cell lines and target variables are displayed in Table VI and referred to as robust feature sets in the following.

Table VII shows the resulting distribution of the absolute percentage error for the selected robust feature sets. C^{ap}, e.g., may be determined for HT-29/B6, IPEC-J2, and MDCK I with a worst-case deviation of 20%. For 11 out of 21 models, maximum APE values are less than 50 percent. For five models, maximum APE values are less than 25 percent. It can be seen, though, that even with robust feature sets, very high maximum deviations may still occur for individual target variables. For example, the determination of R^{ap} of IPEC-J2 has an average deviation of 4.8%, but in the worst case a deviation of 303.0% (cf. Table VII). As percentiles ($P_{95\%}$) indicate, however, such high prediction errors occur very rarely.

V. Postprocessing using Least Squares Approach

After completing the previous step, all components of an assumed equivalent electric circuit are determined by machine learning with the respective accuracy (section IV). Having obtained estimates for all circuit parameters, however, one also obtains the possibility of synthesizing impedance spectra for the given values of the circuit parameters. This is the basis for the postprocessing step that we propose here.

Analogously to the generation of training data (section III), the determined parameter values are used to model two synthetic impedance spectra for control conditions and Nystatin application. Both newly synthesized spectra are then compared to the originally observed spectra. The least-squares method is then employed to adjust the parameter values so that the synthesized pair of spectra matches the original.

A. Approach

Using the machine learning predictions, a new pair of impedance spectra $\hat{Z} = (\hat{Z}_1, \ldots, \hat{Z}_{84})^\top$ is calculated using equation 2. Note, that one does not necessarily obtain an equivalent circuit whose corresponding pair of spectra matches the original observed pair of spectra $Z = (Z_1, \ldots, Z_{84})^\top$, as described in the next Chapter. The Residual Sum of Squares (*RSS*) is a suitable measure for this purpose, which is calculated according to the following formula:

$$RSS(Z, \hat{Z}) = \sum_{i=1}^{84} (Z_i - \hat{Z}_i)^2 \qquad (22)$$

It is desirable that the predicted equivalent circuit exhibit similar electrical behavior to the measured cell, i.e., have a minimum *RSS*. For this purpose, the nonlinear least-squares curve fit method from the Python library SciPy was applied [45]. The *Trust Region Reflective* variant was chosen because of its capability to specify upper and lower bounds for each parameter when calling the method. The predictions of the machine learning models with Robust Feature Sets serve as initial values for the optimization from which a local minimum of the *RSS* is searched for. The bounds from Table II were used and the measurement points were weighted with the standard deviations from the error modeling of section III-B.

Note, that both improvement and deterioration of the values are possible. In previous attempts, no improvement of R^{sub} could be achieved. For this reason and because the predictions of the machine learning models were already sufficiently good, R^{sub} was fixed to its initial value.

B. Results

We first present results on how well the *RSS* could be minimized by the least-squares approach. Secondly, we evaluate relative deviations (APE) of the equivalent circuit parameters.

1) Reduction of the Residual Sum of Squares: For all test samples, the solver was able to obtain parameter values with a lower *RSS* compared to the initial values of the machine learning. For all cell lines, the average reduction of *RSS* by optimization is greater than 99%. (cf. Figure 5). An example of a pair of impedance spectra before and after post-processing is shown in Figure 6. Looking at the 84 complex measuring points, the average *RSME* could be reduced from 20.2 $\Omega \cdot cm^2$ to 2.0 $\Omega \cdot cm^2$ for HT-29/B6, from 67.6 $\Omega \cdot cm^2$ to 2.6 $\Omega \cdot cm^2$ for IPEC-J2 and from 61.1 $\Omega \cdot cm^2$ to 2.5 $\Omega \cdot cm^2$ for MDCK I.
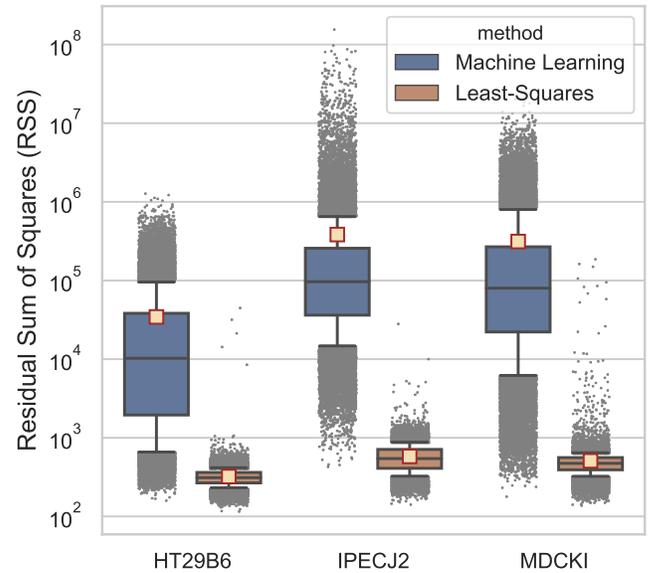

Figure 5. Residual Sum of Squares (*RSS*) of predicted impedance spectra pair before and after least-squares postprocessing as box plots with arithmetic mean (□) and whiskers showing the 10th and 90th percentiles.

2) Circuit parameters: Table VIII shows a detailed overview of derived *APE* values after least-squares postprocessing. Note, that the deviations for R^{sub} do not change because the parameter was fixed during execution. For the targets R_2^{ap}, R^{para} and R^{ap}, a significant improvement is observed for all cell lines. The largest improvement is seen for R_2^{ap}, where, for example, for MDCK I the *MAPE* is reduced from 11.8% to 0.4% and also the maximum *APE* is 99.7% instead of 215.2%. It becomes clear that post-processing delivers better results especially when good initial values are already available, whereas hardly any improvements can be found for the worst-case samples with maximum *APE*. This can be seen in the statistical distribution of *APE*, where, for example, for C^{bl} the median improves for all cell lines, whereas the maximum *APE* increases. The median *APE* worsens in only one case, namely in the determination of R^{bl} for HT/29-B6 by 0.1%. Otherwise, clear improvements can be observed.

TABLE VIII. ABSOLUTE PERCENTAGE ERROR AFTER POSTPROCESSING

	HT-29/B6				IPEC-J2				MDCK I			
	mean	*med*	$P_{95\%}$	*max*	*mean*	*med*	$P_{95\%}$	*max*	*mean*	*med*	$P_{95\%}$	*max*
C^{ap}	0.3	0.2	1.0	9.5	0.3	0.1	1.0	17.7	0.3	0.1	1.2	18.3
C^{bl}	1.3	0.8	4.2	45.1	0.6	0.2	2.4	50.7	0.7	0.2	2.7	44.3
R^{sub}	1.3	1.0	3.7	9.9	1.3	0.9	4.1	20.0	2.3	1.7	6.4	20.7
R^{ap}	4.2	1.8	17.2	76.2	3.0	0.7	12.0	309.7	3.4	0.8	14.6	343.1
R^{bl}	2.2	1.1	7.9	36.6	1.3	0.3	5.3	98.8	1.4	0.3	6.4	94.3
R^{para}	1.7	0.7	6.7	38.3	1.0	0.3	3.6	84.5	1.7	0.4	6.9	120.5
R_2^{ap}	0.9	0.5	3.1	99.6	0.4	0.2	1.2	212.9	0.4	0.3	1.2	99.7

TABLE IX. FEATURE SUBSET IMPORTANCE FOR PREDICTIONS WITH RANDOM FOREST AND $I^3_{\perp(Y)}$

Features	C^{bl}			R^{bl}			R^{para}			R^{sub}		
	HT-29/B6	IPEC-J2	MDCK I	HT-29/B6	IPEC-J2	MDCK I	HT-29/B6	IPEC-J2	MDCK I	HT-29/B6	IPEC-J2	MDCK I
$S_{\Re(Y)} \times S_{\Im(Y)}$	1.3	1.6	2.8	2.5	0.7	3.1	78.6	68.6	73.1	0.2	0.2	1.6
$N_{\Re(Y)} \times N_{\Im(Y)}$	6.5	4.8	24.5	5.3	1.4	4.7	1.2	1.3	3.2	0.2	0.2	0.6
$D_{\Re(Y)} \times D_{\Im(Y)}$	42.9	35.0	23.3	69.4	87.7	78.9	1.9	2.5	4.1	0.1	0.1	0.4
$R_{\Re(Y)} \times R_{\Im(Y)}$	26.5	5.6	7.3	7.6	3.4	2.1	4.4	5.1	11.1	0.1	0.1	0.4
$\mathrm{stat}\left(S_{\Re(Y)}\right) \times \mathrm{stat}\left(S_{\Im(Y)}\right)$	2.3	2.0	1.4	1.3	2.0	1.7	10.2	19.5	3.2	93.3	88.3	10.4
$\mathrm{stat}\left(N_{\Re(Y)}\right) \times \mathrm{stat}\left(N_{\Im(Y)}\right)$	1.3	47.9	37.6	1.5	0.3	1.3	0.4	0.5	0.8	6.0	11.1	86.3
$\mathrm{stat}\left(D_{\Re(Y)}\right) \times \mathrm{stat}\left(D_{\Im(Y)}\right)$	8.5	1.8	1.4	9.6	1.7	6.5	0.7	0.6	1.0	0.0	0.0	0.2
$\mathrm{stat}\left(R_{\Re(Y)}\right) \times \mathrm{stat}\left(R_{\Im(Y)}\right)$	10.8	1.3	1.7	2.8	2.8	1.8	2.7	1.9	3.5	0.0	0.0	0.2

VI. Discussion

A. Relevance for Biomedical Applications

Impedance spectroscopic techniques are increasingly gaining importance in biomedical applications like monitoring the growth of cultured epithelial and endothelial cells (e.g., retinal pigment epithelium, gastrointestinal tract cells, pulmonary cells, blood-brain-barrier models [46][47]), or alterations of barrier function during pharmacological studies [48][49][46]. Furthermore, impedance spectroscopy is the only technique that allows functional distinction between epithelial and subepithelial properties of ex vivo tissue, such as intestinal biopsies of patients with suspected barrier impairment.

Reliable, automated determination of epithelial properties is becoming increasingly important, as, during the past decade, impedance spectroscopy of epithelia has started to move from pure basic research or investigations on biopsies to clinical applications in patients. Several groups have developed various endoscopic devices that allow in vivo impedance measurements, e.g., in the GI tract of critically ill patients that may develop mucosal ischemia [50] or in the esophagus from patients suffering from gastroesophageal reflux disease [51][52][53].

If the technique is to be used on a routine basis, however, reliable automatization for the evaluation of impedance spectra is indispensable. On one hand, manual evaluation of impedance spectra to extract the physiologically relevant parameters requires extensive user training and is time-consuming, as individual spectra need to be fitted by complex equations [24]. On the other hand, currently available systems usually only record R^{trans} (i.e., the sum of the subepithelial and epithelial resistance) or even only relative alterations in R^{trans} over the time-course of an experiment, as estimation of absolute R^{trans} values is too error-prone [54][46].

B. Combined Consideration of Two-State Measurements

The accurate determination of apical and basolateral properties with one-path impedance spectroscopy has been considered an unsolved problem in clinical physiology. In this study, we have been able to demonstrate that this issue can be solved by observing the cell under two different conditions, or states, namely before and after the apical addition of Nystatin. It became clear that machine learning techniques are well suited to process pairs of impedance spectra, with a significant improvement compared to determination based on a single measurement under control conditions alone (see Table V). First attempts to construct new features from two different spectra were successful and showed that this step improves the determination of epithelial properties by machine learning.

C. Feature Extraction

From the results shown in section IV, we can draw some preliminary conclusions regarding the usefulness of the extracted features and feature sets.

1) Representations: In the context of this study, none of the selected representations showed to be consistently superior. However, among the regression models with the smallest MAPE, algebraic forms of impedance and admittance are noticeably more frequent than polar forms (cf. Table V). The average improvement of the MAPE by the right choice of representation is about 0.9 percentage points. Therefore, we assume that the representation does not play a major role in the method presented here.

2) Feature Importance: Scikit-learn's Random Tree Regressor provides the output of feature importances. The value of feature importance lies in the interval [0; 1] and is the larger, the greater the influence on the regression. Notice that the sum of all feature importances of an input feature set is equal to 1. The input feature set can be divided into subsets and compared to each other by summing the individual importances. Table IX shows the feature subset importances for the random forest regressions of C^{bl}, R^{bl}, R^{para} and R^{sub} using $I^3_{\perp(Y)}$ as input.

3) Statistical Features: The extraction of statistical features has worked well in previous work related to epithelial impedance analysis, which is why the approach was retained for this work [1]. The benefit of using statistical features varies for different targets. In Table IX, e.g., it can clearly be seen that the statistical features are of high importance in predicting R^{sub} and C^{bl} through random forest. This is particularly evident in the prediction of R^{sub}. Here, with a summed feature importance of $\geq 97\%$, decisions are made almost entirely based on the statistical features. For C^{bl} of IPEC-J2 and MDCK I, the statistical features in $I^3_{\perp(Y)}$ obtain high importance and could contribute to a $0.9 - 1.9$ percentage point improvement in prediction (cf. Table Vb). Because of the ease of calculation and the possible positive effect on prediction, the use of descriptive statistics for feature extraction in the analysis of impedance spectra can be further recommended.
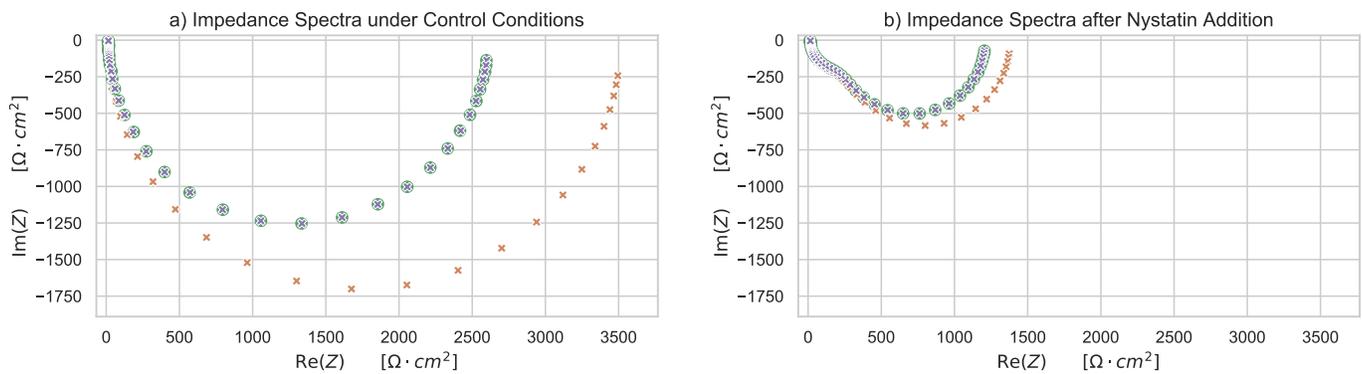
Figure 6. Pair of impedance spectra before and after Nystatin addition: Original measurement from data set (●), pair of spectra derived from Machine Learning predictions (✖), pair of spectra derived from predictions after least-squares postprocessing (✖)

4) Differences and Ratios: The frequency-wise calculation of the differences and ratios from the pairs of spectra is a completely new approach resulting from the two-state consideration. Looking at the feature importance from Table IX, it becomes very clear that the features extracted in this way work very well for transcellular targets such as C^{bl} and R^{bl} with an feature importance between 31% and 91%. On the other hand, for the quantities that do not belong to the transcellular route, i.e., R^{para} and R^{sub}, this approach is of little use. Differences and ratios not only yielded high feature importance for random forests but were also able to contribute to a $1.8-2.8$ percentage point improvement in prediction (see Table Ve). Additionally, more complex features could be extracted from the two-state consideration to further improve prediction.

D. Machine Learning and Least-Squares Optimization

In this work, machine learning methods following statistical and neuro-inspired paradigms were used for the pairwise analysis of epithelial impedance spectra. In particular, the *MAPE* values from Table V show that with the right combination of feature set and machine learning model, very good predictions of all sought targets are possible. With a maximum training time of a regression model of a few hours on a 4-core CPU, the overall training period was acceptable despite the large sample size. Therefore, no further time measurements were taken during training. A drawback of the presented study is the brute-force approach in the search for optimal feature sets. Here, an elaborated search for optimal feature sets would have saved training effort, for example in the unprofitable use of different impedance representations.

This computational study was able to show the promising influence of least-squares postprocessing for machine-learning-based analysis. Figure 6 shows the original impedance spectra of a measurement from the data set and the spectra derived from the machine learning predictions with and without least-squares optimization. It can be seen that the machine learning predictions without least-squares optimization may result in impedance spectra that are significantly different from the measured ones. In this respect, posterior optimization using a least-squares approach has been shown to be a suitable approach to better align a predicted pair of curves to the observed ones. The greatest empirical evidence represents the average reduction of the *RSS* by more than 99% for all

cell lines (see Figure 5). Due to the much better alignment, the equivalent circuit obtained after optimization is much better suited to describe the electrical behavior of the cell. Moreover, with the help of the least-squares optimization, an improvement in the predictions of R^{para}, R^{ap} and R^{ap}_2 by several percentage points could be achieved.

Least-squares methods, as used in our postprocessing, usually require suitable initial values from which a local minimum is found. The choice of suitable initial values for optimization problems is a frequent topic in various disciplines [55][56]. In the field of epithelial impedance spectroscopy, the use of randomly chosen initial values may lead to poor convergence behavior. Manual initialization by an expert, on the other hand, offers no possibility of automation. The presented method could fill this gap by using machine learning to find suitable initial values in an automated way. Accordingly, e.g., in the field of crystal lattice determination, machine learning techniques have recently been used to find suitable initial values for subsequent optimization [57].

E. Limitations

A notable limitation of the presented method is the occurrence of highly increased prediction errors in a few cases, especially for R^{ap} with a maximal relative deviation of more than 300% for MDCK I and IPEC-J2 (see Table VIII). Because the apical resistance decreases after apical addition of Nystatin, R^{ap} is the only parameter that changes during the experimental setup, while the others remain the same. For a reliable approximation of R^{ap}, basolateral addition of Nystatin might be useful to keep R^{ap} constant during both measurements.

Looking at the absolute percentage error plotted against the Nystatin factor d_{nyst}, we see for all targets and cell lines that the highly increased prediction errors occur only for $d_{nyst} < 0.4$. In Figure 7, this can be seen for R^{ap}, R^{bl}, and R^{para} as examples. Note, that in the given experimental setup, the amount of Nystatin added can freely be chosen. Based on this, we filtered and re-evaluated the test data set for samples with $d_{nyst} \geq 0.4$ (see Table X). For the reduced test data set, the overall performance of the proposed method is significantly better with a maximum APE of 36.5%. However, the assumption $d_{nyst} \geq 0.4$ is not applicable to the general case, because smaller d_{nyst} values, derived from a substantial
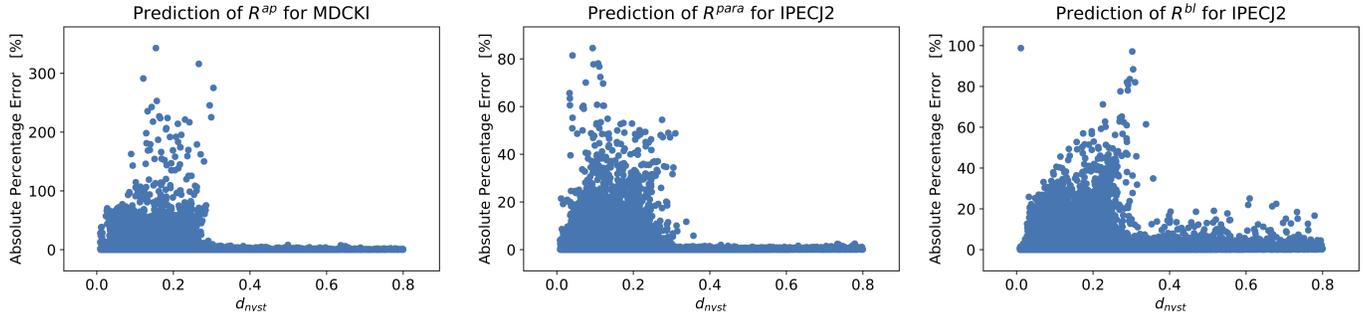
Figure 7. Absolute percentage error plotted against the Nystatin factor d_{nyst}.

addition of Nystatin, are often required to distinguish apical and basolateral membranes in the impedance curve.

As the presented method to determine electric properties of epithelia involves modeled impedance spectra, applicability is naturally depending on the modeling and sampling of the data. The previously modeled data that is used here is in good accordance with data obtained from measurements on epithelial tissue [25]. While a variety of cell lines and functional states is considered in our work, results and estimations obtained here are still limited to the specific scenarios modeled. Also, characteristics of the training and test data influence characteristics of estimation methods. For example, the precision of supervised learning methods tends to decrease if the number of samples decreases. To this end, e.g., the distribution of target values needs to be considered (cf. Figure 4).

VII. Conclusions and Future Work

While impedance spectroscopy is a convenient measurement technique, quantifying the electric bipolarity of epithelial tissue with traditional approaches is challenging. For some parameters, it may not only be error-prone, but practically impossible with reasonable efforts. In this computational study, we have shown that this aim may be achieved with good results by combining machine learning regression with a least squares-based postprocessing step. Key concepts were the combination of measurements obtained under two different tissue conditions as well as a systematic feature extraction approach. Due to detailed and realistic modeling, we suggest this approach is valid for the epithelial cell lines HT-29/B6, IPEC-J2 and MDCK I under control conditions, as well as under the influence of Nystatin and EGTA.

All in all, the method outlined and evaluated in the present study represents a significant progress in order to achieve a routine evaluation of impedance spectra obtained in a clinical measurement setting. Several challenges are envisioned, however, that need to be adressed in future work:

1) Compared to impedance spectra recorded from cultured cell layers, data obtained from patients will be subject to a considerably larger scatter within one impedance spectrum, as electrical shielding is limited.

2) There will be a substantial variability between individual spectra, due to variation in the positioning of the electrodes.

3) Larger artifacts have to be recognized, e.g., caused by movements of the patient.

4) In contrast to the GI tract, esophageal epithelia are stratified. Thus, valid equivalent circuits for stratified epithelia have to be developed and tested, before data can be modeled accurately.

A promising strategy is to solve these open questions in our current approach would be the application of generative machine learning methods, which are designed to rebuild characteristics of given training data. Although techniques such as Generative Adversarial Networks (GAN) were originally designed to mimic image data [58], we see much potential in employing them to model spectral data. In future work, we will therefore investigate the usage of such state-of-the-art generative machine learning methods for advanced and highly automated modeling of impedance spectra.

Appendix

A. Prediction for samples with $d_{nyst} \geq 0.4$

In the discussion section, we have pointed out some notable effects of the Nystatin application on the reliability of predictions. As can be seen in Figure 7, high prediction errors occur exclusively in samples with a small Nystatin factor d_{nyst}. Therefore, the test samples were filtered by $d_{nyst} \geq 0.4$ and re-evaluated. From the original 50,000 samples, we obtained reduced data sets with 11043 Samples for HT29B6, 3526 for IPECJ-2, and 2448 for MDCK I. The results of the evaluation are shown in Table X.

TABLE X. ABSOLUTE PERCENTAGE ERROR AFTER POSTPROCESSING FOR SAMPLES WITH $d_{nyst} \geq 0.4$

	HT-29/B6				IPEC-J2				MDCK I			
	mean	*med*	$P_{95\%}$	*max*	*mean*	*med*	$P_{95\%}$	*max*	*mean*	*med*	$P_{95\%}$	*max*
C^{ap}	0.2	0.1	0.5	4.7	0.2	0.1	0.6	4.6	0.1	0.1	0.4	2.4
C^{bl}	1.4	1.0	4.2	45.1	0.8	0.4	2.7	21.2	0.6	0.4	1.7	9.7
R^{sub}	1.4	1.0	3.9	8.3	1.5	1.0	4.5	18.2	2.3	1.7	6.9	14.9
R^{ap}	1.8	1.0	6.0	36.5	0.6	0.3	2.2	13.0	0.4	0.3	1.4	8.1
R^{bl}	2.4	1.4	7.8	33.0	1.0	0.5	3.7	25.1	0.8	0.5	2.5	17.6
R^{para}	0.9	0.5	2.9	15.0	0.2	0.2	0.7	2.5	0.3	0.2	0.9	3.1
R_2^{ap}	1.0	0.6	3.6	17.3	0.3	0.2	1.2	8.8	0.2	0.2	0.7	5.0

References

[1] T. Schmid, D. Günzel, and M. Bogdan, "Automated quantification of the resistance of epithelial cell layers from an impedance spectrum," in Proceedings of the Tenth International Conference on Bioinformatics, Biocomputational Systems and Biotechnologies, 2018, pp. 8—13.

[2] V. Singh et al., "Translating molecular physiology of intestinal transport into pharmacologic treatment of diarrhea: stimulation of na+ absorption," Clinical Gastroenterology and Hepatology, vol. 12, no. 1, 2014, pp. 27–31.

[3] S. Priyamvada et al., "Mechanisms underlying dysregulation of electrolyte absorption in inflammatory bowel disease–associated diarrhea," Inflammatory bowel diseases, vol. 21, no. 12, 2015, pp. 2926–2935.

[4] J. M. Diamond, "Twenty-first bowditch lecture. the epithelial junction: bridge, gate, and fence," Physiologist, vol. 20, no. 1, Feb 1977, pp. 10–18.

[5] M. Furuse et al., "Occludin: a novel integral membrane protein localizing at tight junctions," Journal of Cell Biology, vol. 123, no. 6, 1993, pp. 1777–1788.

[6] J. Ikenouchi et al., "Tricellulin constitutes a novel barrier at tricellular contacts of epithelial cells," Journal of Cell Biology, vol. 171, no. 6, Dec 2005, pp. 939–945.

[7] M. Furuse, H. Sasaki, K. Fujimoto, and S. Tsukita, "A single gene product, claudin-1 or -2, reconstitutes tight junction strands and recruits occludin in fibroblasts," The Journal of Cell Biology, vol. 143, no. 2, 1998, pp. 391–401.

[8] M. Furuse, H. Sasaki, and S. Tsukita, "Manner of interaction of heterogeneous claudin species within and between tight junction strands," The Journal of Cell Biology, vol. 147, no. 4, 1999, pp. 891–903.

[9] S. M. Krug et al., "Charge-selective claudin channels," Annals of the New York Academy of Sciences, vol. 1257, 2012, pp. 20–28.

[10] J. Luettig, R. Rosenthal, C. Barmeyer, and J. Schulzke, "Claudin-2 as a mediator of leaky gut barrier during intestinal inflammation," Tissue Barriers, vol. 3, no. 1-2, 2015, p. e977176, pMID: 25838982.

[11] C. Barmeyer, J. D. Schulzke, and M. Fromm, "Claudin-related intestinal diseases," Seminars in Cell & Developmental Biology, vol. 42, 2015, pp. 30 – 38, claudins Time, Space and the Vertebrate Body Axis.

[12] T. Schmid, D. Günzel, and M. Bogdan, "Automated quantification of the capacitance of epithelial cell layers from an impedance spectrum," in Proceedings of the 7th International Conference on Bioinformatics, Biocomputational Systems and Biotechnologies, 2015, pp. 27–32.

[13] T. Schmid, M. Bogdan, and D. Günzel, "Discerning apical and basolateral properties of HT-29/B6 and IPEC-J2 cell layers by impedance spectroscopy, mathematical modeling and machine learning," PLOS ONE, vol. 8, no. 7, 2013, p. e62913.

[14] B. Jovov, N. Wills, and S. Lewis, "A spectroscopic method for assessing confluence of epithelial cell cultures," American Journal of Physiology, vol. 261, no. 6 Pt 1, 1991, pp. C1196–203.

[15] J. R. Macdonald and W. B. Johnson, Fundamentals of Impedance Spectroscopy. John Wiley & Sons, Inc., 2005, pp. 1–26.

[16] K. S. Cole, "Electric phase angle of cell membranes," The Journal of general physiology, vol. 15, no. 6, 1932, p. 641.

[17] T. Teorell, "Application of "square wave analysis" to bioelectric studies1," Acta Physiologica Scandinavica, vol. 12, no. 2-3, 1946, pp. 235–254.

[18] E. Schifferdecker and E. Frömter, "The ac impedance of necturus gallbladder epithelium," Pflügers Archiv, vol. 377, no. 2, 1978, pp. 125–133.

[19] G. Kottra and E. Frömter, "Rapid determination of intraepithelial resistance barriers by alternating current spectroscopy. i. experimental procedures," Pflügers Archiv, vol. 402, no. 4, 1984, pp. 409–420.

[20] M. Fromm, C. E. Palant, C. J. Bentzel, and U. Hegel, "Protamine reversibly decreases paracellular cation permeability in necturus gallbladder," The Journal of membrane biology, vol. 87, no. 2, 1985, pp. 141–150.

[21] W. Van Driessche et al., "Interrelations/cross talk between transcellular transport function and paracellular tight junctional properties in lung epithelial and endothelial barriers," American Journal of Physiology-Lung Cellular and Molecular Physiology, vol. 293, no. 3, 2007, pp. L520–L524.

[22] D. Günzel et al., "From TER to trans-and paracellular resistance: lessons from impedance spectroscopy," Annals of the New York Academy of Sciences, vol. 1257, no. 1, 2012, pp. 142–151.

[23] S. M. Krug, M. Fromm, and D. Günzel, "Two-path impedance spectroscopy for measuring paracellular and transcellular epithelial resistance." Biophys J, vol. 97, no. 8, 2009, pp. 2202–2211.

[24] T. Schmid, D. Günzel, and M. Bogdan, "Efficient prediction of x-axis intercepts of discrete impedance spectra," in Proceedings of the 21st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN), 2013, pp. 185–190.

[25] T. Schmid, "Automatisierte Analyse von Impedanzspektren mittels konstruktivistischen maschinellen Lernens," Ph.D. dissertation, Universität Leipzig, Germany, 2018.

[26] M. E. Orazem and B. Tribollet, Eds., Electrochemical Impedance Spectroscopy. Wiley, 2008, ch. Methods for Representing Impedance, pp. 309–331.

[27] K. Suzuki, G. Kottra, L. Kampmann, and E. Frömter, "Square wave pulse analysis of cellular and paracellular conductance pathways in necturus gallbladder epithelium," Pflügers Archiv, vol. 394, 1982, pp. 302–312.

[28] C. Bertrand, D. Durand, G. Saidel, C. Laboisse, and U. Hopfer, "System for dynamic measurements of membrane capacitance in intact epithelial monolayers," Biophysical Journal, vol. 75, no. 6, 1998, pp. 2743–2756.

[29] W.-M. Weber, H. Cuppens, J.-J. Cassiman, W. Clauss, and W. Van Driessche, "Capacitance measurements reveal different pathways for the activation of cftr," Pflügers Archiv, vol. 438, no. 4, 1999, pp. 561–569.

[30] I. El Naqa and M. J. Murphy, "What is machine learning?" in Machine Learning in Radiation Oncology. Springer, 2015, pp. 3–11.

[31] A. Krogh, "What are artificial neural networks?" Nature Biotechnology, vol. 26, no. 2, Feb 2008, pp. 195–197.

[32] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," Neural networks, vol. 2, no. 5, 1989, pp. 359–366.

[33] J. R. Quinlan, "Induction of decision trees," Machine learning, vol. 1, no. 1, 1986, pp. 81–106.

[34] L. Breiman, "Random forests," Machine learning, vol. 45, no. 1, 2001, pp. 5–32.

[35] H. Liu and H. Motoda, Feature Extraction, Construction and Selection: A Data Mining Perspective. Springer Science & Business Media, 1998, vol. 453.

[36] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," Journal of Machine Learning Research, vol. 3, 2003, pp. 1157–1182.

[37] J. Kittler, "Feature selection and extraction," Handbook of Pattern Recognition and Image Processing, 1986.

[38] T. Schmid, D. Günzel, and M. Bogdan, "Automated quantification of the relation between resistor-capacitor subcircuits from an impedance spectrum," in Proceedings of the International Joint Conference on Biomedical Engineering Systems and Technologies, ser. BIOSTEC 2014, 2014, p. 141–148.

[39] T. Schmid, D. Günzel, and M. Bogdan, "Automated quantification of the relation between resistor-capacitor subcircuits from an impedance spectrum," in Proceedings of the International Conference on Bio-Inspired Systems and Signal Processing, 2014, pp. 141–148.

[40] B. Schindler, D. Günzel, and T. Schmid, "Synthesized Impedance Spectra Measurements of Epithelial Tissue," Nov. 2021. [Online]. Available: https://doi.org/10.5281/zenodo.5718939

[41] T. Schmid, "Automatisierte analyse von impedanzspektren mittels konstruktivistischen maschinellen lernens," Ph.D. dissertation, Leipzig, 2018.

[42] S. Lewis, D. Eaton, C. Clausen, and J. Diamond, "Nystatin as a probe for investigating the electrical properties of a tight epithelium," The Journal of general physiology, vol. 70, 11 1977, pp. 427–40.

[43] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," International Conference on Learning Representations, 12 2014.

[44] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, 2011, pp. 2825–2830.

[45] P. Virtanen et al., "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," Nature Methods, 2020.

[46] B. Srinivasan et al., "TEER measurement techniques for in vitro barrier model systems," Journal of Laboratory Automation, vol. 20, no. 2, 2015, pp. 107–126.

[47] N. Onnela et al., "Electric impedance of human embryonic stem cell-derived retinal pigment epithelium," Medical & Biological Engineering & Computing, vol. 50, no. 2, 2012, pp. 107–116.

[48] F. A. Atienzar, H. Gerets, K. Tilmant, G. Toussaint, and S. Dhalluin, "Evaluation of impedance-based label-free technology as a tool for pharmacology and toxicology investigations," Biosensors, vol. 3, no. 1, 2013, pp. 132–156.

[49] S. Ramasamy, D. Bennet, and S. Kim, "Drug and bioactive molecule screening based on a bioelectrical impedance cell culture platform," International Journal of Nanomedicine, vol. 9, no. 1, 2014, pp. 5789–5809.

[50] N. E. Beltran and E. Sacristan, "Gastrointestinal ischemia monitoring through impedance spectroscopy as a tool for the management of the critically ill," Experimental biology and medicine, vol. 240, no. 7, 2015, pp. 835–845.

[51] D. Sifrim and K. Blondeau, "Technology insight: the role of impedance testing for esophageal disorders," Nature Clinical Practice Gastroen-terology & Hepatology, vol. 3, no. 4, 2006, pp. 210–219.

[52] F. Ates et al., "Mucosal impedance discriminates gerd from non-gerd conditions," Gastroenterology, vol. 148, no. 2, 2015, pp. 334–343.

[53] T. Matsumura et al., "Evaluation of esophageal mucosal integrity in patients with gastroesophageal reflux disease," Digestion, vol. 97, no. 1, 2018, pp. 31–37.

[54] Z. Maherally et al., "Real-time acquisition of transendothelial electrical resistance in an all-human, in vitro, 3-dimensional, blood–brain barrier model exemplifies tight-junction integrity," The FASEB Journal, vol. 32, no. 1, 2018, pp. 168–182.

[55] P. Tsiotras, E. Bakolas, and Y. Zhao, "Initial guess generation for aircraft landing trajectory optimization," 08 2011.

[56] Q. Yao and H. Tong, "Quantifying the influence of initial values on non-linear prediction," Journal of the Royal Statistical Society: Series B (Methodological), vol. 56, no. 4, 1994, pp. 701–725.

[57] K. Kaufmann et al., "Crystal symmetry determination in electron diffraction using machine learning," Science, vol. 367, 2020, pp. 564–568.

[58] L. Wang, W. Chen, W. Yang, F. Bi, and F. R. Yu, "A state-of-the-art review on image synthesis with generative adversarial networks," IEEE Access, vol. 8, 2020, pp. 63 514–63 537.