analytical chemistry

ARTICLE

pubs.acs.org/ac

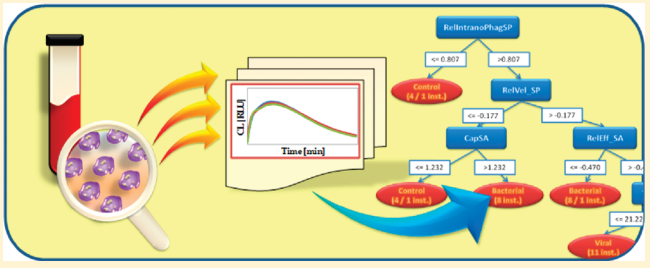
Differentiation between Viral and Bacterial Acute Infections Using Chemiluminescent Signatures of Circulating Phagocytes

Daria Prilutsky,[†,‡,§] Evgeni Shneider,[∥] Alex Shefer,[⊥] Boris Rogachev,[¶] Leslie Lobel,[†] Mark Last,[‡] and Robert S. Marks*[,§,#]

[†]Department of Virology, Faculty of Health Sciences, [‡]Department of Information Systems Engineering, Faculty of Engineering Sciences, [§]National Institute of Biotechnology in the Negev, [∥]Emergency Department, [⊥]Division of Internal Medicine, Department of Medicine H, [¶]Department of Nephrology, Soroka Medical Center, and [#]Department of Biotechnology Engineering, Faculty of Engineering Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel

Supporting Information

ABSTRACT: Oftentimes the etiological diagnostic differentiation between viral and bacterial infections is problematic, while clinical management decisions need to be made promptly upon admission. Thus, alternative rapid and sensitive diagnostic approaches need to be developed. Polymorphonuclear leukocytes (PMNs) or phagocytes act as major players in the defense response of the host during an episode of infection, and thereby undergo functional changes that differ according to the infections. PMNs functional activity can be characterized by quantification and localization of respiratory burst production and assessed by chemiluminescent (CL) byproduct reaction. We have assessed the functional states of PMNs of patients with acute infections in a luminol-amplified whole blood system using the component CL approach. In this study, blood was drawn from 69 patients with fever (>38 °C), and diagnosed as mainly viral or bacterial infections in origin. Data mining algorithms (C4.5, Support Vector Machines (SVM) and Naïve Bayes) were used to induce classification models to distinguish between clinical groups. The model with the best predictive accuracy was induced using C4.5 algorithm, resulting in 94.7% accuracy on the training set and 88.9% accuracy on the testing set. The method demonstrated a high predictive diagnostic value and may assist the clinician one day in the distinction between viral and bacterial infections and the choice of proper medication.



Despite medical advances, differentiation between a viral and bacterial etiology based on clinical symptoms is currently very difficult, being at the same time of utmost importance to clinicians.[1,2] Accurate and rapid information in this regard can result in commencement of appropriate treatment for a viral or bacterial disease: the former infection generally resolves with supportive care only, and the latter usually necessitates the administration of antibiotics. Treating viral infections with antibiotics is ineffective and contributes to the development of antibiotic resistance, allergic reactions, toxicity and greater healthcare costs.[2−4] The most common ways to diagnose bacterial infections currently are microbiological cultures, elevated leukocyte and neutrophil counts, C-reactive protein level, and erythrocyte sedimentation rate.[5] A viral infection diagnosis includes tests for viral antigens, PCR and determination of antibody titers, as well as normal or low white blood cells count. Those techniques are time-consuming and sometimes have poor sensitivity and specificity. Therefore, an alternative rapid, sensitive, and commercially available diagnostic tool for determining whether the infection is viral or bacterial in origin is needed.

Polymorphonuclear leukocytes (PMNs) or phagocytes, often elevated in blood, play a major role in the defense response of the host during an episode of infection.[6,7] These cells inherit different pieces of information in the case of different etiological agents stimulating immunity. The molecular environment in a specific infection is pretuning the "warrior" immune cells, thus, preparing them for future tasks and acting as a disease marker.[8] Upon activation of phagocytes, an increase in the consumption of molecular oxygen occurs, which results in the production of reactive oxygen species (ROS), a process collectively called respiratory burst.[9−11] When luminol is added to the system, the ROS generation is accompanied by light emission (chemiluminescence, CL). Thus, CL is a sensitive marker of the oxidative potential of phagocytes, which can be recorded as a luminol-dependent CL (LCL).[11,12]

Attempts have been made to correlate the primed activity of circulating phagocytes with the severity of disease and its outcome.[13−18] The correlation of phagocytic CL activity to peritoneal dialysis associated clinical states was previously established and the C4.5 decision-tree algorithm demonstrated an

Received: March 14, 2011
Accepted: April 26, 2011
Published: April 26, 2011

84.6% predictive accuracy.[14] The phagocytic function resulting from viral exposure was studied previously, including the resulting CL response produced as a result of viral infection.[19−24] A number of reports have also shown altered metabolic activity in PMNs in response to bacterial infection.[19,20,25] It was found that bacterial infections generally elevate host cellular CL.[26] The hypothesis of the present study is that phagocytes are of high predictive value for use as a sensitive infection marker for distinguishing between different kinds of infections providing us with a timely diagnostic tool.

Herein, we characterized the functional states of phagocytic cells from patients with acute infections with different etiologies by applying the aforementioned approach of component chemiluminescence sensing in a luminol-amplified whole blood system. We applied different data mining algorithms in order to discover specific patterns for each type of infection. The highest classification accuracy was reached by the C4.5 algorithm,[27] which also provides the most interpretable model in the form of a decision tree. Afterward, the rules were used to classify blind cases, achieving 88.9% prediction accuracy.

To conclude, the approach of inspecting whole-blood CL as a measure of respiratory burst, analysis of kinetics, and classification of clinical groups using data mining algorithms demonstrated a high predictive diagnostic value and may assist in proper treatment modalities a very short time after admission of the patient to the hospital.

EXPERIMENTAL SECTION

Study Population. Sixty-nine patients were recruited from Emergency Room and Internal Medicine Departments, Soroka University Medical Center. These cases were recruited on the basis of appearance of high fever (>38 °C) and basic signs of possible acute infection in the moment of admission to the hospital. The patients recruited formed two groups: one group with identified bacterial infection (33 instances) and the other group with identified viral infections (or nonbacterial infections; 36 instances). The instances were randomly divided into two data sets: 51 instances as a training set (to induce the model) and 18 as a blind set (to evaluate the predictive performance of the model). The bacterial infections were mainly from two groups: urinary tract infections (UTI, 5 instances) and respiratory tract infections (28 instances). Diagnosis was based on signs and symptoms of high grade fever (>38 °C), cough, dyspnea, auscultatory findings, typical findings of infiltrates in chest X-rays, and positive microbiological urinary or/and blood culture. Cases from the bacterial group suffered from infections caused by a single or multiple bacteria (UTI caused by *Escherichia coli*, *Pseudomonas aeruginosa*, *Enterococcus*). Viral infections caused by viruses (Human Rhinovirus (HRV, common cold), Influenza virus (flu) or swine flu (H1N1), Para influenza virus) were diagnosed by clinical observations of fever, cough, running nose, headache, and myalgia, while others (H1N1) using PCR. The viral diagnoses were supported by a negative microbiological culture, lack of findings in chest X-rays and either normal or slightly elevated leukocyte counts. It was assumed that the main source of illness was bacterial or viral, while mixed infections were eliminated for better clarification between the border between viral and bacterial etiology. For control cases, we measured and recorded blood CL kinetics of 6 healthy subjects without clinical symptoms of infection. All control instances were used in the training set. A detailed list of all the patients recruited and their affiliation to a specific clinical group are found in the Supporting Information (SI), Table S-1. The study protocol was approved by the Helsinki committee of Soroka University Medical Center (Certification Nums. 4241, 4599, 4888). All patients gave their informed consent for this study.

Reagents. Zymosan-A (Sigma Chemical Co., Z4250), used as a stimulating agent, was opsonized for 30 min at 37 °C in each corresponding patient plasma sample (10 mg/mL) and prepared in Krebs-Ringer phosphate buffer prior to use[13] (20 mg/mL). Luminol (Bio-Rad HRP substrate kit 170-5040) was used to amplify chemiluminescence activity. In some experiments, fMLP (ICN Biomedicals, 1511170) was used (10^{-9} M final concentration) for priming of the phagocytes.

Chemiluminescence Assays. Peripheral whole blood (1:100 (v/v) final dilution) was used to avoid the appearance of artifacts due to the isolation of PMNs and to preserve conditions that are close to the *in vivo* cellular environment. Samples, at a total volume of 200 μL, contained whole fresh blood diluted in KRP, luminol, and zymosan in KRP as was described before.[14] After addition of opsonized zymosan, the sample contents were mixed and CL was measured. Three CL systems were investigated: standard, priming, and aging as was described before[14] and mentioned in Supporting Information.

Data Analysis. Data analysis was performed in the following steps: (1) the experimental CL curves were recorded; (2) full sets of kinetic parameters were calculated based on the decomposition of the experimental CL curves to the three biological components; (3) patients with similar clinical states were placed into defined groups (such as group with bacterial infections, group with viral infections and healthy controls); (4) data mining algorithms were applied to induce classification models for discriminating between clinical groups; (5) the algorithm predictive performance was evaluated.

The recorded kinetic CL curves were presented as a sum of three biological components representing the following processes:[13,28] (1) extracellular ROS generation connected to phagocytosis; (2) intracellular ROS generation, connected to phagocytosis; (3) intracellular ROS, not connected to phagocytosis. To decompose the chemiluminescent curve into the three aforementioned components, the PeakFit program was employed, using Poisson-type distribution equations for each component,[29] considering the boundary conditions for the time appearance of each peak: first component t_1 [1−6] min, second component t_2 [7−13], third component t_3 [>13] min.[30−32] The first and the third components were decomposed arbitrarily to two additional components to represent the CL multimodal curve in more details:

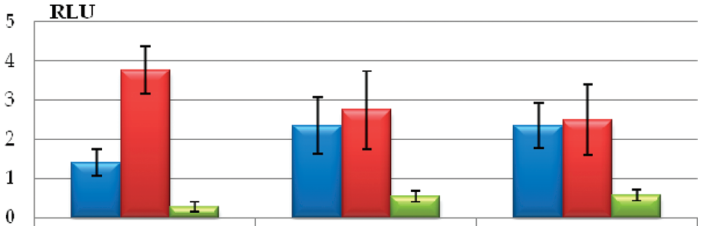
$$t_1\ [1-6]\ \text{min} = t_{1_1}[1-2]\ \text{min} + t_{1_2}\ [3-6]\ \text{min}$$

$$t_3\ [>13]\ \text{min} = t_{3_1}\ [13-25]\ \text{min} + t_{3_2}\ [>26]\ \text{min}$$

As a result, the final number of curve components was five. The detailed decomposition of one of the curves is represented in the Supporting Information, Figure S-1. Matching the sum of the CL components obtained by modeling to that of the experimental curve was achieved by determining the minimum sum of the squared differences.[30−32] For each patient, 82 kinetic parameters derived from three systems (standard, priming, and aging systems) were calculated, while the kinetic parameters were published before.[14] The parameters calculated for each specific curve are indicated in Table S-2 of the Supporting Information.

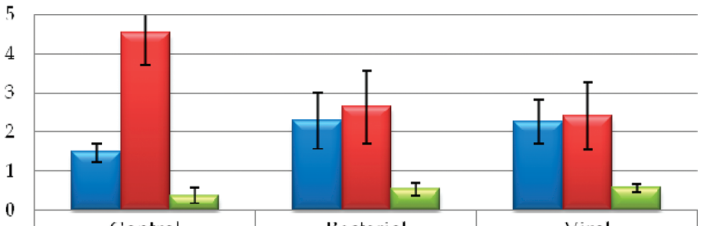
Models Induction. The 75 cases were used as a final data set. Three data mining algorithms (C4.5 (J48), SVM and Naive Bayes)

(A) Standard system



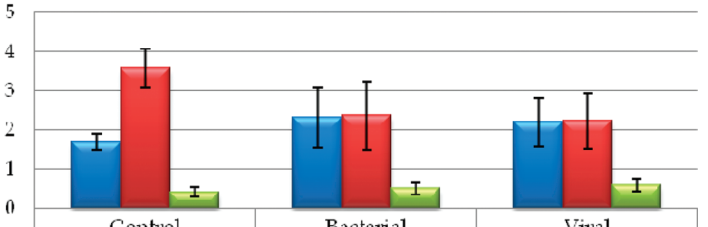
	Control	Bacterial	Viral
Capacity	1.407	2.344	2.347
Effectivness	3.776	2.737	2.499
Velocity	0.280	0.552	0.571

(B) Priming system



	Control	Bacterial	Viral
Capacity	1.480	2.299	2.272
Effectivness	4.561	2.648	2.405
Velocity	0.380	0.537	0.579

(C) Aging system



	Control	Bacterial	Viral
Capacity	1.692	2.320	2.203
Effectivness	3.568	2.369	2.228
Velocity	0.418	0.513	0.592

Figure 1. Description and comparison of derived respiratory burst parameters (capacity, effectiveness, and velocity) in three clinical states (bacterial infection, viral infection, and control) in the standard (A), priming (B), and aging (C) systems. Functional states of PMNs are assigned according to the calculated parameters as "resting" state in the case of control and "effective" or "fighting" state in both types of infections.

were applied to induce classification models for differentiation of clinical groups using RapidMiner 4.3 open-source data mining tool (http://rapid-i.com) and Weka 3.5.6 Machine Learning Software.[33] Ten-fold stratified cross-validation was applied 10 times using a different local random seed in each iteration. The results are averaged for the 10 iterations of 10-fold cross validation. In parallel, the data set was split randomly into training and blind sets (24% of the instances). All classification results are presented in a 95% confidence interval. To improve classification performance, feature selection was applied as a preprocessing step. The wrapper approach was used with forward selection feature generation scheme implemented by the feature selection operator in the RapidMiner software.

RESULTS AND DISCUSSION

Whole Blood Chemiluminescence. Patients recruited were divided into two main groups, namely, those resulting from bacterial or viral etiology, and these were compared to the control group in the studies conducted. UTI cases were not excluded from this study and were included in the clinical group with bacterial infections, because of the similar CL pattern to respiratory tract infections.

Owing to the experimentation performed, the patients with different infections could be grouped according to their chemiluminescent behavior and the functional states of phagocytes could be characterized. Figure 1 presents the main parameters

dx.doi.org/10.1021/ac200596f |*Anal. Chem.* 2011, 83, 4258–4265

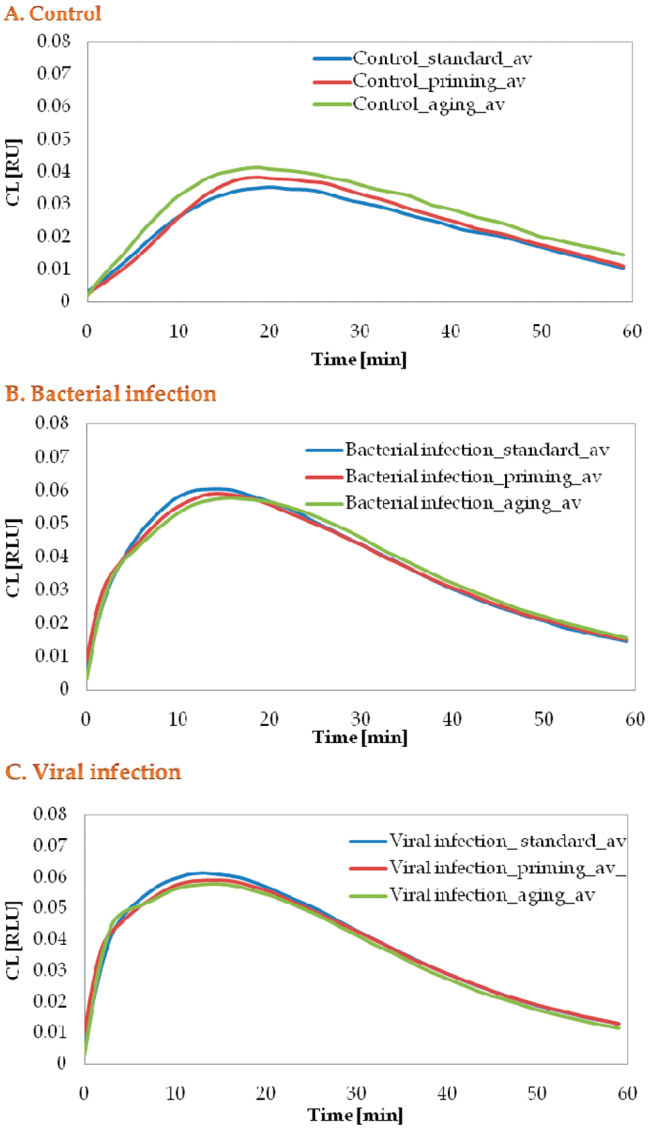


Figure 2. CL pattern data recorded using standard, priming, and aging systems and averaged by different groups. Three clinical states are represented: control (A), bacterial infection (B), and viral infection (C). Different CL patterns are used for distinguishing between different acute clinical conditions of patients by their distinct patterns of PMNs functionality.

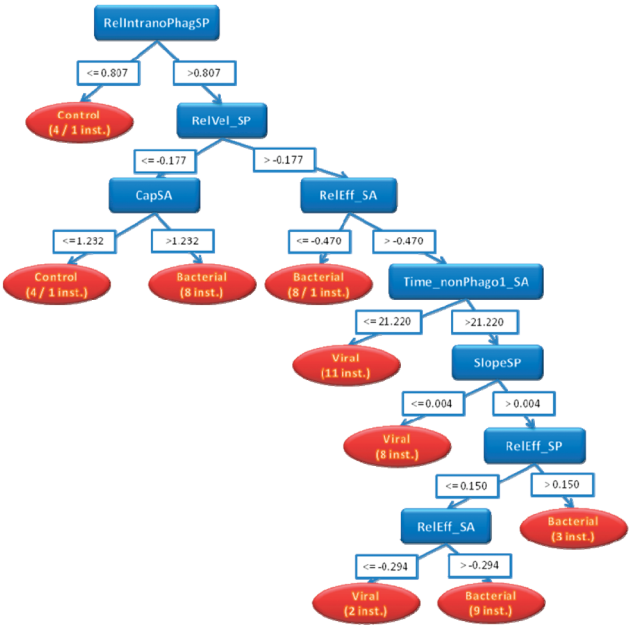
(capacity, effectiveness, and velocity) that are used in the estimation of functionality of the immune system and Figure 2 represents the CL curves in case of different infections. The data was analyzed and compared to healthy, noninfected controls demonstrating the earlier described "resting" pattern ($C_LE_HV_L$ = $Capacity_{Low}Effectiveness_{High}Velocity_{Low}$).[14,34] The resting pattern is characterized with a low capacity to generate ROS, while phagocytosis of the prey is performed with high effectiveness, expressing the healthy state of the cells; and low velocity, as the cells are not activated and not ready for the defense and killing. In comparison to controls, it can be observed that the two infected groups have higher capacity and velocity in the standard system, revealed by the presence of infection, leading to an elevated cellular CL response (Figure 1A). The effectiveness of the phagocytosis in both infections is lower than in the control cases, reflecting some of the ROS generation processes that flow outside the cells.

The CL kinetics of *bacterial infection* are associated with high capacity, lower effectiveness than in controls, and high velocity in the standard system (Figures 1A and 2B), corresponding to an "effective" ($C_HE_HV_H$), very close to the "fighting" ($C_HE_LV_H$) functional state of the phagocytes. This state corresponds to the activated state of the immune system in case of infection (as it was already shown before[14,34]): the phagocytes have high capacity to generate ROS, which occurs predominantly inside the cells. This oxidative potential of cells is achieved by processes directly associated with phagocytosis. As can be seen examined by kinetics parameters and CL curves, priming and aging influenced the pattern of the curve. Effectiveness in priming and aging systems was lower, as it is also represented by a high initial slope of the CL curve and, as a result, highest extracellular ROS generation. The velocity of the priming and aging curves is higher, and the peak appears earlier in time, representing activation of phagocytes.

The CL kinetics of *viral infection* cases is very similar to bacterial infections, exhibiting activation of the system as a result of stimulation of the immune system during infection. The viral cases showed high capacity, effectiveness lower than controls, and high velocity in the standard system (Figures 1A and 2C), defined, as in the case of bacterial infection, as an "effective" or "fighting" functional state of the phagocytes. In comparison to bacterial infection, effectiveness is lower in case of viral infections, as seen in the kinetics curves presented in Figure 2, where the initial slope of the viral group's kinetics is much sharper than in case of bacterial infection. As revealed by the kinetics parameters, priming had little influence on the pattern of the curve. In the priming system, capacity, effectiveness, and velocity remain almost the same as compared to the standard system (Figures 1B and 2C). In the aging system, capacity is lower, as revealed in the aging kinetic curve that is slightly lower than the two other curves. Effectiveness is slightly lower because of the more emphasized extracellular part of the kinetics (Figures 1C and 2C). Velocity is higher, representing the dominant extracellular part and the ROS generated mainly in the processes connected to phagocytosis.

In the case of bacterial infections, the system showed high capacity to produce ROS and lower effectiveness as compared to the controls, as well as high velocity, meaning that the ROS generation occurs mainly by processes directly associated with phagocytosis and has substantial extracellular origin. In general, the functional state of bacterial infection corresponds to the "effective" or "fighting" functional state. As reported in previous studies,[19,20,26] PMNs in the majority of patients with active bacterial infection are found in an activated state, both functionally and metabolically, and these produce increased CL. The results of the present study support these previous reports, while also showing an increased CL response with high capacity and velocity. Moreover, our research shows that the main part of ROS production during phagocytosis takes place in the intracellular space.

In case of a viral infection, the system showed almost the same behavior as in the bacterial cases, except for the predominant extracellular origin of the ROS generated in phagocytosis, which is higher in viral infections, resulting in lower effectiveness. As reported in previous studies, influenza viral infection has been associated with decreased phagocytic cell function and inhibition of chemiluminescent responses of PMNs to various stimuli,[24] together with other studies that report an induction in phagocytic function.[21−23] The results presented in our study address viral infections in a nonseparated manner as predominantly respiratory



Classified as →	Bacterial infection	Viral infection	Control	*Total*	*Percent correct*
Bacterial infection	24	0	1	**25**	**96%**
Viral infection	1	24	1	**26**	**92.3%**
Control	0	0	6	**6**	**100%**

Testing set: Classification of 18 Blind Cases with Infections According to the Induced Model

Classified as →	Bacterial infection	Viral infection	*Total*	*Percent correct*
Bacterial infection	6	2	**8**	**75**
Viral infection	0	10	**10**	**100**

Detailed Accuracy by Class: Blind Cases with Bacterial or Viral Infections

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
Bacterial	0.75	0	1	0.75	0.857	0.875
Viral	1	0.25	0.833	1	0.909	0.875

Figure 3. Decision tree and confusion matrix of patients in the training and testing sets. Three clinical groups were classified with 94.7% accuracy in a training set: 24 out of 25 cases in group with bacterial infections (96%), 24 out of 26 cases in group with viral infections (92.3%), and 6 cases of control (100%) were classified correctly. The tree is based on 7 parameters chosen by the algorithm from the 82 parameters that were calculated according to the component approach. Sixteen out of 18 blind cases (88.9%) were classified correctly according to the induced model, with detailed accuracy represented by class.

infections (i.e., influenza and HRV-mediated infections were considered together and taken as an average of the different viruses). Thus, on average, the phagocytes are induced and the CL curve is elevated as compared to the control. The reason for the lower effectiveness in the case of viral infection can be explained by taking into account the different natures of viral infections, with not only the influenza-causing virus being present. Furthermore, as shown in previous studies,[35] HRV produces oxidative burst mainly outside the cell, thus, lowering the effectiveness.

To conclude, the two infections exhibit completely different types of chemiluminescent signatures, each one described by its own characteristics, proving that the oxidative potential of the innate immunity can indeed serve as a prognostic or diagnostic marker. Moreover, we show herein that the infectious states do show an effective type of phagocytic behavior, as supported by the behavior of peritonitis[14] and pulmonary abscess cases.[34]

Phagocyte CL Information Can Differentiate between Clinical Groups. The kinetic parameters previously described were calculated in all clinical cases, and the data set was checked to evaluate to what extent the proposed methodology can serve as a relevant tool to describe clinically primed and challenged phagocyte diversity. The induced classification models resulted in a statistical separation of three clinical groups: bacterial infections, viral infections, and controls. The most accurate and interpretable model was induced by the C4.5 algorithm as compared to Support Vector Machines (SVM) and Naive Bayes (Table S-3) in training/testing configuration. In the case of SVM, the linear model was selected to be the most accurate. Naive Bayes Classifier resulted in 73.7% prediction accuracy in the training set (similarly to linear SVM). In Figure 3, we present the most accurate induced model, which is the decision tree constructed by C4.5. It was obtained using the value of a minimum of 2 instances per leaf. In the training set, three clinical groups were classified with 94.7% accuracy.

Control Group. In differentiating the control group, three parameters were chosen by classification task: the ratio of the intracellular nonphagocytosis capacity of the primed sample divided by the intracellular nonphagocytosis capacity of the standard sample (RelIntranoPhagSP), the ratio of velocity of primed sample to the velocity of standard sample (RelVel_SP), and capacity of the aging system (CapSA). Relative velocity in the priming system should be lower in control cases when comparing to all viral infections and to part of the bacterial infections. This is supported by Figure 1B, where velocity of

controls is lower than in case of any infection. The capacity of the aging system should be lower than that found in bacterial cases. This finding, derived from the classification model, is supported by Figure 1C, where the capacity of control is much lower than in both types of infection.

Bacterial Infections Group. There are four clusters or leaves formed to differentiate this clinical group. The first parameter that is substantial in differentiating bacterial cases from controls is CapSA, capacity of aging system. This parameter is estimated to be higher in bacterial infections than controls, which is supported by Figure 1C. Second parameter, which appears twice as a node in a decision tree, is RelEff_SA, relative effectiveness of aged system as compared to the standard system. According to the model presented, RelEff_SA should be higher in part of the cases with bacterial infection, when compared to viral infections. This finding is supported by the parameters derived from CL kinetics (Figure 1C), where the effectiveness of the aged system in bacterial infections is higher than in viral infections. An additional parameter that plays a role is the RelEff_SP, relative effectiveness in primed system as compared to the standard system. This parameter should be higher in part of the cases with bacterial infections, as compared to cases with viral infections, which is supported by Figure 1C.

Viral Infections Group. Four main parameters influence the differentiation of the viral infections group, which is divided to 3 clusters during classification task. The first is the Time_nonPhago1_SA (time to peak of the last portion of nonphagocytosis related CL of the aging system) which is chosen for differentiation of the main part of the clinical cases with viral infections. In these infections, the time peak for production of ROS in not connected to the phagocytic processes chosen to be earlier in time as compared to bacterial infections, thus, reflecting faster kinetics. This is supported by the chemiluminescent parameters of aged kinetics presented in Figure 1C, where velocity is the highest in cases with viral infections. The second parameter differentiating the viral group is SlopeSP, that is, the ratio of the peak magnitude to time required to reach this point, which should be low as compared to some of the cases with bacterial infections, thus, reflecting slow kinetics in the primed system. This finding cannot be seen on the CL parameters, mainly because the parameters are averaged for all cases, while the decision tree differentiates on a case-by-case basis. Another two parameters are relative effectiveness of primed (RelEff_SP) and aged (RelEff_SA) systems, which should be lower in viral infections when compared to bacterial ones, seen in Figure 1, panels B and C, respectively.

To summarize, the parameters which successfully distinguish between the clusters matched the chemiluminescent parameters of the systems presented in Figure 1, thus, revealing the patterns unique for each group. The oxidative potential of phagocytes, expressed by CL kinetics in various infections, has enabled the differentiation using the classification model, correlating well with the evaluation of the functional state of the fighting cells. Control cases were characterized with low capacity when compared to the infected ones, as it was defined by the model and supported by the kinetics parameters. The model illustrated that the ROS generation is more effective in cases of bacterial infections, when compared to viral infections, showing that the ROS is mainly generated intracellularly in bacterial phagocytosis. The CL curves demonstrated a higher extracellular part in the viral group via a sharper initial slope, thus, expressing low effectiveness. The kinetics in viral cases was fast, as it is portrayed in the model and in the CL parameters (Figure 1) as well. Therefore, the combination of modeling and the molecular description of the systems was successful in defining and describing each clinical group: the groups with infections showed a high activation of the immunity, which was expressed as a high velocity of ROS generation, and fast phagocytosis; high effectiveness to produce ROS for defense, but still lower than in the case of a potential fight in healthy individuals; and high capacity of oxidative potential.

Evaluation of Blind Cases. On the basis of the decision tree model in Figure 3, blind cases can be evaluated and their group membership determined. As such, 18 blind cases were examined, applying the classification rules to the phagocyte function of those patients and assigning them to the most appropriate group. Sixteen of the 18 cases were correctly classified (88.9%), with 75% of the bacterial infections and 100% of viral infections being classified correctly using C4.5 algorithm (Figure 3).

In the case of the SVM using linear model (polynomial kernel with degree of kernel = 1) and Naive Bayes Classifier, three clinical groups were classified with 73.7% in the blind set. Despite the limited amount of data used in this research, the approach showed very good compliance and high prediction accuracy in classification of the blind cases. According to the obtained results, the presented method has a high probability to become a valid diagnostic tool.

Comparison between Algorithms. Data mining tools can be very useful to organize and find specific patterns and parameters that are special to some types of infections in the data sets presented. Together with that, some questions should be answered: Which tools can be optimal? What would be the best possible model? Does this model correlate with biological significance and does it provide insight into molecular events associated with phagocytic processes and ROS generation? Comparison of the prediction accuracies of differentiation to clinical groups derived by various classification algorithms (based on 10-times 10-fold stratified cross-validation) is presented in Table S-3. Without feature selection, C4.5 resulted in 50.8% ± 3.9% prediction accuracy, SVM achieved a 50.8% ± 1.2% accuracy, and Naïve Bayes Classifier resulted in 47.6% ± 1.7% prediction accuracy.

Feature selection has substantially improved the prediction accuracy after 10-fold cross-validation in all algorithms: C4.5 accuracy was improved from 50.8% ± 3.9% to 69.2% ± 2.5% in case of 2 instances per leaf; SVM accuracy was improved from 50.8% ± 1.2% to 62.4% ± 2.5%; Naive Bayes accuracy was improved from 47.6% ± 1.7% to 66.8% ± 1.4%. All differences were found to be statistically significant ($p < 0.001$) using *t*-test. The best performance was achieved in case of C4.5, as in the case of independent training/testing sets. The decision tree classifier may be actually the most useful while comparing to other algorithms in presenting classification problems, because of the "white-box" straightforward model with easy-to-read explicit classification rules that can be interpreted biologically, characterizing each specific clinical group.

CONCLUSIONS

Bacteria and viruses are the most frequent agents causing acute respiratory infections, such as pneumonia and viral bronchitis. Bacterial cultures and immune-fluorescence tests, as well as PCR for viral antigens, are used to differentiate bacterial from viral infections in clinics. However, the results of bacterial cultures

often require 24−48 h, and indeed, tests for bacterial and viral etiology may not be available in some clinical settings.[36] Moreover, negative bacterial culture cannot always exclude the bacterial origin of an infection. Unfortunately, there are no highly sensitive validated clinical prediction rules for distinguishing bacterial from viral etiology (with the exception of incomplete, rarely used and complicated PCR technologies), despite the fact that such criteria can guide decisions as to whether to use antibiotics. Limiting the use of unnecessary antibiotics may help prevent the development of antibiotic resistance, reduce the number of patients with adverse effects of antibiotics, and substantially decrease healthcare costs.[36,37] Crucial understanding of early dynamics of infection, interaction between the immune system and the pathogens and, as a consequence, a proper selection of early intervention was the motivation for this study.

It is well-known that phagocytic cells are the first line of defense in our body, and are influenced differently by different clinical set-ups. In our previous studies, we have shown that phagocytes inherit molecular information expressed as extents of ROS generation which is specific to various clinical states in intra-abdominal pathological processes afflicting peritoneal dialysis patients,[14] as well as various infectious diseases (unpublished data). Therefore, the objective of this study was to develop and validate a clinical diagnostic rule based on the inherited information of an innate immunity to differentiate bacterial from viral etiology, using component CL approach and multiple derived kinetic parameters.

Bacterial or viral-infected patients showed completely different CL patterns from healthy controls. As already stated, healthy controls showed a "resting" pattern. In the case of bacterial and viral infections, it can be seen that the CL response was elevated in the capacity of phagocytes to produce ROS and in their velocity to react to the stimulus. In other words, phagocytes in infectious cases were activated as compared to controls, and showed an "effective" or "fighting" functional state. Phagocytes influenced by bacterial infections mainly produce ROS intracellularly, while phagocytes influenced by viral infections produce ROS extracellularly to a greater extent than did bacterial infections.

Results from the present study demonstrate that the level of activation of circulating phagocytes can be a sensitive indicator of infection. The derived CL response from a given patient supplied us with enough information so as to place that particular case closer to either clinical group and associate that case with a corresponding prognosis, while enabling us to differentiate the etiologies in a rapid and sensitive manner. The method is time-saving, easy to perform and can be commercially available, thus, having predictive diagnostic value and could be implemented in various medical institutions as an adjunct to clinical decision making after further investigation.

To find an accurate classification model that possibly can be used by clinicians in the future, we applied three data mining algorithms: C4.5 decision tree, SVM, and Naive Bayes Classifier. We have also used feature selection methods to improve the performance and choose the relevant and nonredundant predictive features that present a biological meaning. The results obtained on a blind set show that the decision tree C4.5 classification algorithm demonstrated the best predictive performance, revealing explicitly interpretable classification rules. Using 10-fold cross-validation, C4.5 has also reached the highest accuracy, but it was lower than in the blind set (almost 70% predictive accuracy versus 88.9% testing accuracy in the blind set).

Future basic molecular studies are needed to further characterize the specific CL kinetic patterns generated from the interactions of PMNs with different types of microorganisms. Additional clinical variables can be added to the database as features, for example, parameters form biochemical blood tests, simple blood counts, microbiological cultures, antibodies ratios, and so forth. We may also improve the classification accuracy by increasing the number of features involved and applying active learning techniques like bagging[38] and boosting.[39] The final outcome of the method application to the clinical setups can be a learning diagnostic tool that is based on various data mining algorithms.

■ ASSOCIATED CONTENT

Supporting Information. Additional information as noted in the text. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Tel.:+ 972 8 6477182. Fax: +972 8 6472857. E-mail: rsmarks@bgu.ac.il.

■ REFERENCES

(1) File, T. M. *Lancet* **2003**, *362*, 1991–2001.
(2) Nuutila, J.; Lilius, E. M. *Curr. Opin. Infect. Dis.* **2007**, *20*, 304–310.
(3) Livermore, D. M. *Lancet Infect. Dis.* **2005**, *5*, 450–459.
(4) French, G. L. *Adv. Drug Delivery Rev.* **2005**, *57*, 1514–1527.
(5) Korppi, M.; Heiskanen-Kosma, T.; Leinonen, M. *Eur. Respir. J.* **1997**, *10*, 1125–1129.
(6) Mayer-Scholl, A.; Averhoff, P.; Zychlinsky, A. *Curr. Opin. Microbiol.* **2004**, *7*, 62–66.
(7) Segal, A. W. *Annu. Rev. Immunol.* **2005**, *23*, 197–223.
(8) Van Dyke, K.; Van Dyke, C. *Methods Enzymol.* **1986**, *133*, 493–507.
(9) Magrisso, M.; Marks, R. S.; *Handbook of Biosensors and Biochips*; Marks, R. S., Cullen, D. C., Karube, I., Lowe, C. R., Weetall, H. H., Eds.; John Wiley & Sons: Hoboken, NJ, 2007; pp 511−529
(10) Rosen, G. M.; Pou, S.; Ramos, C. L.; Cohen, M. S.; Britigan, B. E. *FASEB J.* **1995**, *9*, 200–209.
(11) Dahlgren, C.; Karlsson, A. *J. Immunol. Methods* **1999**, *232*, 3–14.
(12) Vilim, V.; Wilhelm, J. *Free Radical Biol. Med.* **1989**, *6*, 623–629.
(13) Magrisso, M.; Etzion, O.; Pilch, G.; Novodvoretz, A.; Perez-Avraham, G.; Schlaeffer, F.; Marks, R. *Biosens. Bioelectron.* **2006**, *21*, 1210–1218.
(14) Prilutsky, D.; Rogachev, B.; Vorobiov, M.; Zlotnik, M.; Last, M.; Lobel, L.; Marks, R. S. *Anal. Chem.* **2008**, *80*, 5131–5138.
(15) Stevens, D. L.; Bryant, A. E.; Huffman, J.; Thompson, K.; Allen, R. C. *J. Infect. Dis.* **1994**, *170*, 1463–1472.
(16) Wakefield, C. H.; Carey, P. D.; Foulds, S.; Monson, J. R.; Guillou, P. J. *Arch. Surg.* **1993**, *128*, 390–395.
(17) Maderazo, E. G.; Woronick, C. L.; Albano, S. D.; Breaux, S. P.; Pock, R. M. *J. Infect. Dis.* **1986**, *154*, 471–477.
(18) Zimmerman, G. A.; Renzetti, A. D.; Hill, H. R. *Am. Rev. Respir. Dis.* **1983**, *127*, 290–300.
(19) Solberg, C. O.; Kalager, T.; Hill, H. R.; Glette, J. *Scand. J. Infect. Dis.* **1982**, *14*, 11–18.
(20) McCarthy, J. P.; Bodroghy, R. S.; Jahrling, P. B.; Sobocinski, P. Z. *Infect. Immun.* **1980**, *30*, 824–831.
(21) Peterhans, E. *Biol. Trace Elem. Res.* **1997**, *56*, 107–116.
(22) Mills, E. L.; Debets-Ossenkopp, Y.; Verbrugh, H. A.; Verhoef, J. *Infect. Immun.* **1981**, *32*, 1200–1205.
(23) Peterhans, E. *Virology* **1980**, *105*, 445–455.

(24) Abramson, J. S.; Lyles, D. S.; Heller, K. A.; Bass, D. A. *Infect. Immun.* **1982**, *37*, 794–799.
(25) Briheim, G.; Stendahl, O.; Dahlgren, C. *J. Infect. Dis.* **1987**, *156*, 676–680.
(26) Barbour, A. G.; Allred, C. D.; Solberg, C. O.; Hill, H. R. *J. Infect. Dis.* **1980**, *141*, 14–26.
(27) Quinlan, J. R. *C4.5: Programs for Machine Learning*; Morgan Kaufmann Publishers Inc.: San Francisco, CA, 1993.
(28) Magrisso, M. J.; Alexandrova, M. L.; Bochev, P. G.; Bechev, B. G.; Markova, V. I.; Benchev, I. C. *J. Biochem. Biophys. Methods* **1995**, *30*, 257–269.
(29) Magrisso, M. J.; Bechev, B. G.; Bochev, P. G.; Markova, V. I.; Alexandrova, M. L. *J. Biolumin. Chemilumin.* **1995**, *10*, 77–84.
(30) Magrisso, M.; Marks, R. Publication Number: WO/2006/092787, International Application No.:PCT/IL2006/000272, 2005.
(31) Magrisso, M.; Marks, R. Publication Number: WO/2006/092788, International Application No.:PCT/IL2006/000273, 2006.
(32) Magrisso, M.; Marks, R. Patent num in submission: 22611-WO-07. 2007.
(33) Witten, I. H.; Frank, E. *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed.; Morgan Kaufmann: San Francisco, CA; 2005.
(34) Magrisso, M. Y.; Alexandrova, M. L.; Markova, V. I.; Bechev, B. G.; Bochev, P. G. *Luminescence* **2000**, *15*, 143–151.
(35) Kaul, P.; Biagioli, M.; Singh, I.; Turner, R. *J. Infect. Dis.* **2000**, *181*, 1885–1890.
(36) Moreno, L.; Krishnan, J. A.; Duran, P.; Ferrero, F. *Pediatr. Pulmonol.* **2006**, *41*, 331–337.
(37) Mulholland, E. K.; Simoes, E. A.; Costales, M. O.; McGrath, E. J.; Manalac, E. M.; Gove, S. *Pediatr. Infect. Dis. J.* **1992**, *11*, 77–81.
(38) Breiman, L. *Machine Learning* **1996**, *24*, 123–140.
(39) Freund, Y.; Schapire, R. In *Proceedings of the Second European Conference on Computational Learning Theory*; Springer-Verlag: London, U.K., 1995.