

Neural Network Classification of Pharmaceutical Active Ingredient from Near Infrared Spectra

Nigel Yee, Ashley Yan,
Electrotechnology Department,
Unitec Institute of Technology,
Private Bag 92025, Victoria Street West,
Auckland 1142, New Zealand.
nyee@unitec.ac.nz, hyan@unitec.ac.nz

Abstract—This paper presents results from a scoping study undertaken with the intention of demonstrating the applicability of reflectance Near-Infrared (NIR) spectroscopy in classification of pharmaceutical type based on active ingredient. The part of the process selected for the scoping study is the packaging step in the manufacturing of pharmaceuticals. The rationale for the selection of the packaging step is in older processing lines the product is classified after tablet coating and before blister packaging using visual automatic inspection techniques however for 100% conformance a more discriminating technique is required. In this study, NIR spectra (with wavelengths between 400nm and 1100nm) were obtained for samples pertaining to 3 different types of pharmaceuticals Quetiapine, Ibuprofen and Paracetamol. The dimensionality of the data set was reduced using Principal Components and the data was fed into a back propagation neural network configured to classify the data based on active ingredient type. The recognition rates achieved in this study were high enough to suggest that NIR spectroscopy is a viable method of ensuring 100% identification of pharmaceutical type prior to packaging for the pharmaceuticals tested.

Keywords—*Neural network; near infrared spectrometer; active ingredient.*

I. INTRODUCTION

Quality control is an essential element of pharmaceutical production processes because of the high safety levels demanded in assuring the pharmaceuticals quality [1] and the legal requirements that finished pharmaceuticals conform to and the number of different measures and tests that are prescribed by various international regulatory regimes [2,3].

Within most pharmaceutical manufacturing lines the quality control tool of automatic inspection prior to packaging has become common place [4] and most modern pharmaceutical lines use advanced inspection systems operating in the visual region of the Electromagnetic Spectrum to assess product shape, colour and pressed label identifier [5]. In general these systems are very good at ensuring the correct product is packaged however automatic inspection systems using only the visual spectrum are limited by their range of wavelengths when tasked with making an inference on the tablets chemical composition [6].

NIR spectroscopy is defined as compilation and analysis of electromagnetic radiation between the wavelengths of 700-2500nm [7]. The NIR electromagnetic spectrum has the capacity to make an inference on a pharmaceutical tablet's chemical composition and provides a rapid analytical technique to acquire product information [8]. NIR spectroscopy has thus become an important tool in product conformance testing and has been widely used in the pharmaceutical sector with successful applications having been reported [9,10].

Implementation of NIR measurement technologies is considered a major cost for many pharmaceutical manufacturers because new machinery equipped with NIR instrumentation is expensive and replacing existing machinery with new NIR instrumentation equipped machinery means realizing the depreciation of the older machinery. A risk associated with deployment of NIR sensing technology is that the NIR spectrometer's configuration may be unsuitable for measurements of the product under investigation due to product size, unique nature of the chemical composition, and spectral range selected [11]. In respect to these uncertainties it is often considered prudent to undertake a pre-deployment scoping study first. In this scoping study the back propagation neural network classifier was selected to determine the predictive power of NIR instrumentation due to it being a general purpose classification tool which is able to model both linear and non-linear relationships [12].

In this scoping study, production tablets at the packaging stage had their spectra captured with a low cost spectrometer (operating over the range of 400-1100nm) mounted on an optical layout table. The pharmaceutical spectra are preprocessed and then a neural network is used to perform a classification; finally recognition statistics are computed using the classification data. The recognition statistics are then used to make a decision about the potential of NIR spectral analysis to recognize pharmaceutical tablet type.

II. METHOD AND MATERIALS

A. Manufacture of Pharmaceuticals

The production process for manufacture of pharmaceutical tablets involves several steps

- In the first step of granulation, a mixture of active ingredient, cellulose, and starch is blended and then granulated in a fluid bed reactor that is supplied with a binder solution and then dried to produce granulate.
- During the second step, granulate is milled to spread the active ingredient homogenously.
- For the third step, granulate is pressed to produce a non-coated tablet.
- The fourth step requires the tablets are coated with a film of lacquer.
- Lastly the tablets are packaged in a high speed packaging line for distribution to human consumers.

These steps are illustrated in a process flow diagram shown in Fig.1 below.

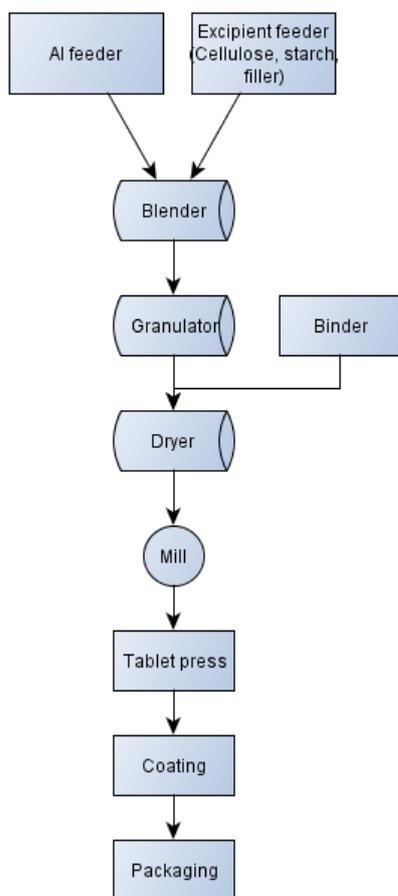


Fig. 1. Process flow diagram of pharmaceutical manufacturing process.

B. Pharmaceutical Formulation

The formulation studied consisted of tablets containing 100mg, 200mg and 300 mg dosages Quetiapine. The Quetiapine tablets had the same concentration of active ingredient but were of different sizes to produce different dosages. Two other tablet types, Ibuprofen (500mg) and Paracetamol (500mg), were used as samples.

C. Recording of NIR Spectra

The pharmaceutical tablet's spectrum was captured using a Zeiss MMS1 VIS-ENH spectrometer [13]. This spectrometer used a fibre optic coupling cable to acquire the signal from the tablets. The spectrometer's electronic interface was manufactured by Australian Innovative Engineering, Queensland [14] and used an Ethernet connection which was interfaced with a computer.

Pharmaceutical tablets samples were placed on a holding unit mounted to an optical layout table (see Fig. 2). The samples were illuminated with a 100W halogen source that was placed 1cm from the sample in the diffuse reflectance configuration.



Fig. 2. Fibre optic coupling input to Zeiss MMS1 spectrometer, sample holder (with table located on it) and lighting configured in diffuse reflectance mode.

D. Preprocessing of the Spectra

The measurements from the spectrometer were taken in reflectance mode and the mathematical treatment of the spectral variables was

$$x_i = \log_{10}(1/R), \quad (1)$$

where R represents the reflectance spectral intensity at a specified wavelength.

A dark reference spectrum from was captured to remove the effects of stray light and a reference spectrum from a ceramic tile was used to standardize the signal to changes in illumination intensity.

$$x_{istd} = (x_i - x_d) / x_c, \quad (2)$$

where x_d is the reference dark spectrum and x_c is the ceramic tile reference spectrum.

The Standard Normal Variate (SNV) algorithm was implemented to reduce the effects of light scatter produced by shape and geometric irregularities in the objects under observation [15]. The SNV works on individual sample spectra by a transformation of the centre of each spectrum and then scaling the spectrum by its own standard deviation.

$$x_{i\ SNV} = (x_{istd} - \bar{x}) / S_d, \quad (3)$$

where S_d represents the spectrum's standard deviation.

E. Principal Component Dimension Reduction

The principal components transformation is a linear geometric transformation that reconfigures the data from a high-dimensional space to a space of fewer dimensions [16]. The principal components algorithm reduces the dimensions of the spectral data matrix by forming basis vectors that

contain information relating to the variation in the spectral data matrix. The basis vectors are in computed as the eigenvectors \mathbf{p} , of,

$$\mathbf{X}\mathbf{X}^T, \quad (4)$$

where \mathbf{X} is the $m \times n$ matrix of spectral intensity measurements containing m observations of n wavelength. In general, the higher order components tend to contain noise and the first few components contain chemical composition information so only the first few a eigenvectors $\mathbf{p}_1, \dots, \mathbf{p}_a$ are retained in the principal component dimension reduction.

F. Neural Network Classification

The neural network used for the classification problem was a back propagation neural network utilising a Jacobian matrix (see Fig. 3). This neural network consisted of three layers: an input layer, an intermediate hidden layer and an output layer and is configured in the feed forward direction [16]. A number of trials were performed using different numbers of nodes within the hidden layer for the purpose identifying a structure that would give an acceptable preliminary result in a reasonable time without undue stress being placed on the computing infrastructure. From these trials a structure containing and input layer that comprised of 5 inputs nodes (one input node per selected principal component), one hidden layer that comprised of 10 nodes and an output layer that comprised of three nodes (one output node for each class of tablet) was used for the purpose of reporting the neural networks classification performance in trial experiments. The classification information from the output layer was used to compute pharmaceutical tablet recognition statistics.

G. Data Analysis

A Hewlet Packard laptop that used two 2.8GHz processors running the Windows 7 operating system and SpecAquisition

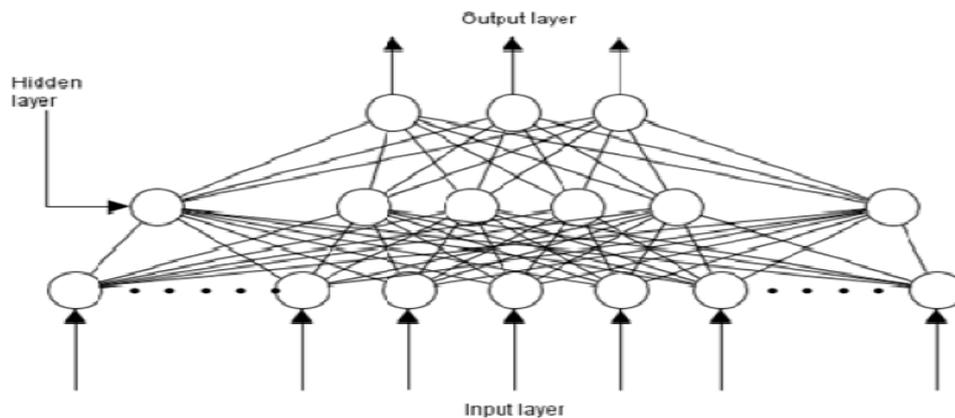


Fig. 3. Neural network with input layer, hidden layer and output layer.

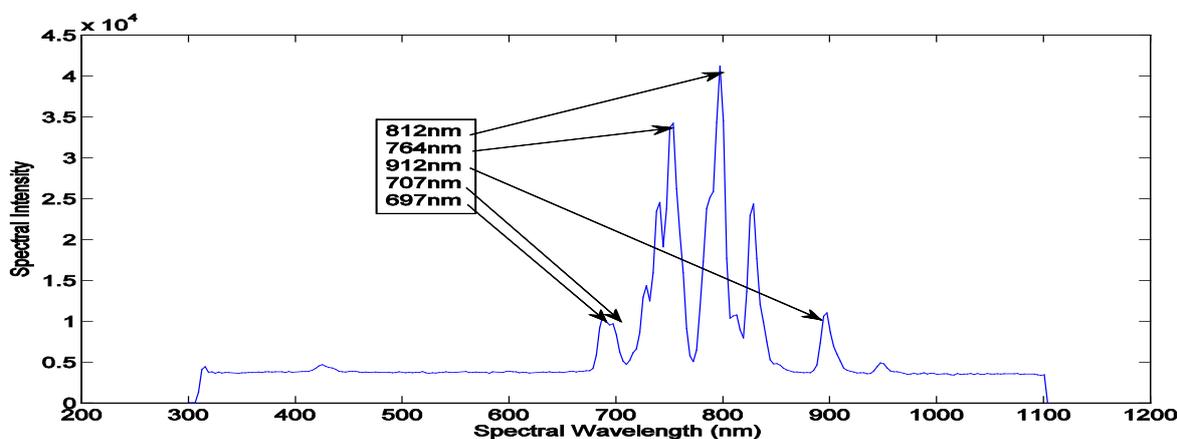


Fig. 4. Plot of spectra from an argon mercury lamp vs. wavelength.

V30 software from Australian Innovative Engineering, Victoria, [14] was used to capture the data from the spectrometer. The data was then saved and imported into Matlab 2011 [18] where the data analysis of standardisation SNV pre-processing, principal component dimension reduction and neural network classification was undertaken.

resolution of the Zeiss MMS1 instrument is 2.7nm but the practical resolution is determined by the optical components used and their resulting optical aberrations. Inspection of two prominent spectral peaks located at 696.5nm, 706.7nm indicate that a resolution of around 10nm is achievable in practice.

III. RESULTS AND DISCUSSION

A. Spectrometer Calibration Results

The spectrometer was calibrated using a mercury argon lamp, spectrums were recorded of excited mercury argon gas and a plot of spectral intensity vs. wavelength for the calibration lamp is shown in Fig. 4. The theoretical spectral

B. Analysis of Principal Components

The standardized SNV treated spectra had principal component dimension reduction applied and then the cumulative variance of the principal components was computed. The cumulative variance statistic revealed that the first factor (PC1) contained most of the chemical and physical information with 89.3% of the dataset's variance.

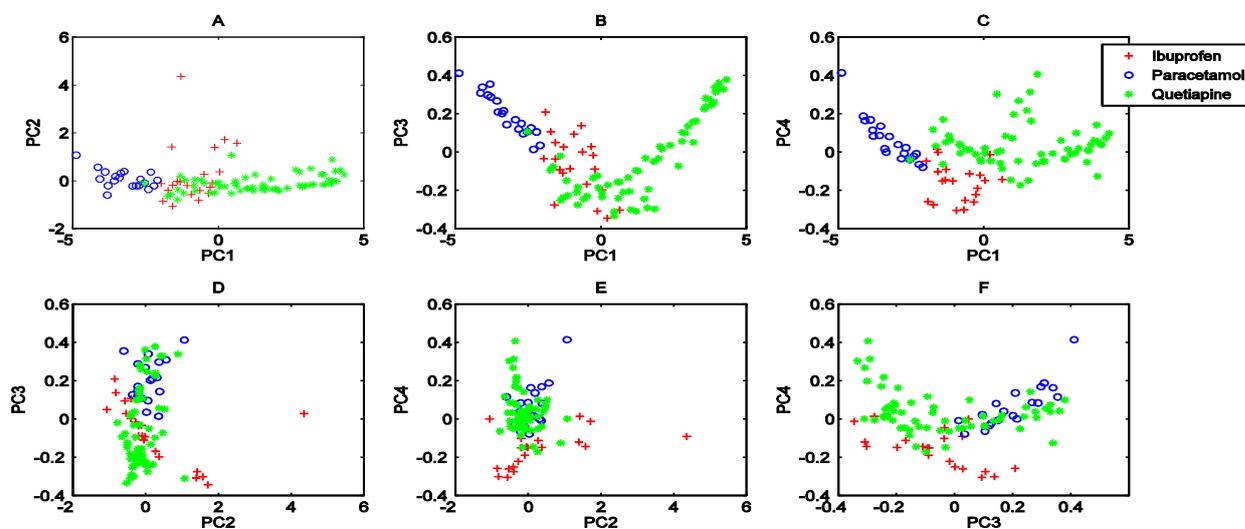


Fig. 5. Principal component scatter plots using the first 4 principal components of the pharmaceutical dataset comprising of 21 Ibuprofen spectrums, 20 Paracetamol spectrums and 61 Quetiapine spectrums.

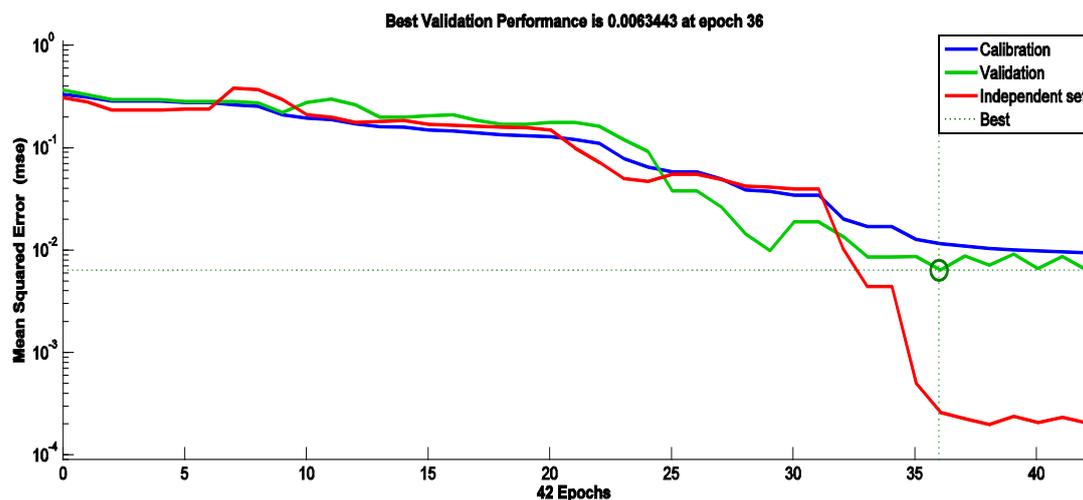


Fig. 6. Neural Network MSE vs. Epoch iterations

Fig. 5 show scatter plots for the first 4 principal components of the pharmaceutical spectra and from inspection of Fig. 5a, 5b and 5c (PC1 vs. PC2, PC1 vs. PC3 and PC1 vs. PC4) a clear trend emerges with Paracetamol samples at different positions between the centre and the left hand side of the plot, Ibuprofen clustered in the centre and Quetiapine at different positions between the centre and the right hand side of the plot. In these plots Quetiapine had the greatest geometric spread of values and this has been attributed to the Quetiapine samples having a greater variability in geometric shape because the Quetiapine samples inspected were of different dosage levels. The remaining plots (Fig 5e, Fig. 5f and Fig. 5g) show much more overlap of the samples and this further suggests that principal component 1 has the greatest discriminatory information

C. Neural Network Training

The neural network was trained using a training set 87 samples (19 Ibuprofen, 15 Paracetamol and 53 Quetiapine) and then an independent set of 15 samples (5 Ibuprofen, 2 Paracetamol and 8 Quetiapine) was used to assess the prediction abilities of the classifier.

The training data set was further broken into two sets, a calibration set of 72 samples and a validation set of 15 samples. During training at 36 epoches, the Mean Squared

Error (MSE) of the validation set produced a minima of 0.0064 before a consistent increase in MSE; indicating termination of training (see Fig. 6). The trained neural network was then used to undertake the classification of samples in the training set and then the prediction set.

D. Recognition Performance of the Neural Network

In reference to Table I, the first recognition statistic ‘*’ refers to the data set of 87 spectra used in training the classifier. The recognition statistic is a percentage value of tablets correctly classified relative to total number of pharmaceutical tablets present for each respective class in the training data set. The second recognition statistic ‘**’ refers to an independent set of 15 spectra not used in training the classifier. The recognition statistic is a percentage value computed from the number of tablets correctly classified relative to total number of tablets present for each respective class in the independent data set.

The classification statistics achieved in recognising the different tablets with the classifier (see Table I) show that the classifier achieved a recognition value of 100% in identifying Ibuprofen and Paracetamol and 98.1% for Quetiapine in the training set. The reverse error was a recognition value of 100% in identification of Ibuprofen and Quetiapine in the training set.

TABLE I. Recognition statistics computed using classification data from Pharmaceutical tablet spectra.

Active ingredient	Number in calibration set	Number correctly classified	* Percentage correctly classified	Number in prediction set	Number correctly classified	** Percentage correctly classified
Ibuprofen	19	19	100	2	2	100
Paracetamol	15	15	100	5	5	100
Quetiapine	53	52	98.1	8	8	100

The inability of the neural network classifier to obtain 100% recognition of Quetiapine does not necessarily suggest that the classifier is inefficient at separation of Quetiapine from the other two tablet types tested. The overall prediction statistics are excellent however the Quetiapine result indicates that the classifier will omit samples of Quetiapine in order to obtain a 100% reverse error for Quetiapine. In a production situation if Quetiapine was being packaged and the intention of the classifier was to ensure only Quetiapine tablets be packaged then this classifier would be considered highly efficient however if the secondary aim was to ensure that all the Quetiapine be recognised and packaged (to optimize production) then this classifier has scope for improvement.

IV. FUTURE WORK

This study indicated that the use of NIR reflectance spectroscopy allows fast non-destructive recognition of active ingredients in the different products investigated in a laboratory environment however in a manufacturing situation where the packaging step in the production line handles much larger numbers of different product batches over the course of a year, a much more substantial calibration would be required before it can be concluded that the technique provides sufficient recognition of all the different batches that are produced. We will attempt to address this problem in future work by increasing the sample size and variety of products used in calibrations and test other classification algorithms.

The current investigations were undertaken using an optical layout table within a controlled laboratory environment; a commercial manufacturing environment contains many more sources of errors such as mechanical vibration, temperature fluctuation and potential movement of the product during inspection. Investigations of these error sources were outside the scope of this study and accordingly our future investigations will attempt to develop calibrations that incorporate these sources of errors.

ACKNOWLEDGMENT

The authors wish to thank Suresh Palapati and Raymond Chiu for technical assistance and Unitech Institute of Technology, New Zealand for laboratory spacing.

REFERENCES

- [1] X. Y. Lawrence, "Pharmaceutical quality by design: product and process development, understanding, and control," *Pharmaceutical Research*, vol. 25(4), pp 781-791, 2008.
- [2] B. K. Immel, "A brief history of the GMPs for pharmaceuticals," *Pharmaceutical Technology*, vol. 25(7), pp 44-53, 2001.
- [3] H. Lofgren, and R. de Boer, "Pharmaceuticals in Australia: developments in regulation and governance," *Social science & Medicine*, vol. 58(12), pp 2397-2407, 2004.
- [4] J. Derganc, B. Likar, D. Tomažević, and F. Pernuš, "Real-time automated visual inspection of color tablets in pharmaceutical blisters," *Real-Time Imaging*, vol. 9(2), pp 113-124, 2003.
- [5] M., Bukovec, Ž. Špiclin, F. Pernuš, and B. Likar, "Automated visual inspection of imprinted pharmaceutical tablets," *Measurement Science and Technology*, vol. 18(9), pp 2921, 2007.
- [6] K. M. Morisseau, and C. T. Rhodes, "Pharmaceutical uses of near-infrared spectroscopy," *Drug Development and Industrial Pharmacy*, vol. 21(9), pp 1071-1090, 1995.
- [7] D. A. Burns, and E. W. Ciurczak, (Eds.). *Handbook of Near-infrared Analysis*, CRC press, Boca Raton, Florida, 2007.
- [8] M. Blanco, and I. N. I. R. Villarroya, "NIR spectroscopy: a rapid-response analytical tool," *TrAC Trends in Analytical Chemistry*, vol. 21(4), pp 240-250, 2002.
- [9] T. Herkert, H. Prinz, and K. A.Kovar, "One hundred percent online identity check of pharmaceutical products by near-infrared spectroscopy on the packaging line," *European Journal of Pharmaceutics and Biopharmaceutics*, vol. 51(1), pp. 9-16, 2001.
- [10] H. Forcinio, "Pharmaceutical Industry Embraces NIR Technology." *Spectroscopy*. vol. 18(9), pp. 16-19, 2003.
- [11] W. Wu, and D. L. Massart. "Artificial neural networks in classification of NIR spectral data: Selection of the input." *Chemometrics and Intelligent Laboratory Systems*, vol. 35(1), pp. 127-135, 1996.
- [12] M. C. Sarraguça, and J. A. Lopes, "Quality control of pharmaceuticals with NIR: From lab to process line." *Vibrational Spectroscopy*, vol 49(2), pp. 204-210, 2009.
- [13] Carl Zeiss Jena GmbH, "MMS—Monolithic Miniature Spectrometer," 2015, ftp://ftp.nist.gov/pub/physics/lunarproject/References/Spectrometers/ZEISS%20MMSOverview_t5U.pdf
- [14] Anonymous, SPC220 user manual version 1.4, Australian Innovative Engineering, 90 Muirs Lane, Myrmiong VIC 3341, Australia.
- [15] R. J. Barnes, M. S. Dhanoa, and S. J. Lister, "Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra," *Applied spectroscopy*, vol. 43(5), pp. 772-777, 1989.
- [16] I. Jolliffe, *Principal component analysis*, John Wiley & Sons Ltd, 2002.
- [17] H. Rumelhart, G.E. Hinton and R.J. Williams, "Learning internal representations by error propagation," *Nature*, vol. 323, pp 533, 1986.
- [18] MATLAB 2011 and Statistics Toolbox, The MathWorks, Inc., Natick, Massachusetts, United States.