Reprinted from

Symposium on Machine Processing of Remotely Sensed Data

June 21 - 23, 1977

The Laboratory for Applications of
Remote Sensing

Purdue University
West Lafayette
Indiana

IEEE Catalog No.
77CH1218-7 MPRSD

ADVANCEMENTS IN MACHINE-ASSISTED ANALYSIS OF MULTISPECTRAL DATA FOR LAND USE APPLICATIONS

PHILIP H. SWAIN
Purdue University

ABSTRACT

Results are reported of a three-year study participated in by the Laboratory for Applications of Remote Sensing of Purdue University, the Center for Advanced Computation of the University of Illinois, and the Geographic Applications Program of the U. S. Geological Survey. The outcome of the study has been a demonstration of the feasibility of applying digital analysis of satellite data to land use inventory and mapping. Advancements have been made in the areas of data analysis techniques, data processing products, and education and training of personnel within the potential user agency.

I. INTRODUCTION

In 1973, a joint study by the Laboratory for Applications of Remote Sensing (LARS) of Purdue University, the Center for Advanced Computation (CAC) of the University of Illinois, and the Geographic Applications Program (GAP) of the U. S. Geological Survey was initiated for the purpose of assessing the applicability of advanced remote sensing systems to land use classification and analysis. Based on initial successes, the study was expanded, eventually encompassing a three-year period and going beyond simple application of existing technology to advance the state-of-the-science in areas where the need was clear and the possibilities promising. The net outcome of the three-year effort has been a conclusive demonstration of the feasibility of applying digital analysis of satellite data to land use inventory and mapping.[1] Significant contributions have been made to the remote sensing technology in general, particularly with respect to cartographically oriented applications. Personnel within GAP have learned and transferred to their agency the capability to utilize the remote sensing technology as it applies to their mission. A number of large-scale demonstration projects have drawn heavily on the developments begun in conjunction with this project (for examples, see references 2, 3, and 4).

A. OBJECTIVES

Broadly stated, the objectives of the joint study can be categorized as follows.

Data Analysis Techniques. Develop and demonstrate the utility of digital data processing techniques as applied to LANDSAT data for classifying land use and land-use change.

Data Processing Products. Develop and demonstrate the capability to produce maps and tabular results in various scales, formats, etc., and with sufficient quality to meet the requirements of the user community.

Education and Training. Transfer to USGS personnel the capability to utilize the digital data analysis technology in an effective and insightful manner.

B. SUMMARY OF ACCOMPLISHMENTS

Data Analysis Techniques. Where computer-assisted techniques had been applied previously on a very much *ad hoc* basis for land use analysis, this study saw the organization of these techniques into well-defined procedures. The effectiveness of these procedures was demonstrated through application to a wide variety of urban areas. Several digital techniques were specifically developed or substantially refined under this study. These are discussed in somewhat

This work was supported by the U. S. Department of the Interior, Geological Survey, under Contract No. 14-08-0001-14725. The views and conclusions contained in this paper are those of the author and should not be interpreted as necessarily representing the official policies, expressed or implied, of the U. S. Government.

greater detail later in this paper, together with some related research activities.

Data Processing Products. It was demonstrated that commercially available filmwriting systems can produce useful images of the land use classification results. Products from a number of these commercial systems were evaluated, and it was found that the general characteristics, cost, and quality of the products differ substantially. Fortunately, this is an area where considerable development efforts are proceeding on the commercial scene.

Various forms of tabular presentation of the results were also developed. Results were tabulated on the basis of rectangular cells defined by geographical coordinates (e.g., UTM grid cells). Of greater significance, it was demonstrated that arbitrary political or jurisdictional boundaries could be digitized from existing maps, registered to the LANDSAT data and classification results, and the results could then be aggregated on the basis of these political or jurisdictional units. Tabulations of land cover by county were obtained for large portions of Missouri and Indiana, and by census tract for the metropolitan Washington, D.C. area.

Education and Training. A variety of USGS personnel participated in the technique development and data analysis activities and as a result acquired considerable expertise in the application and interpretation of the digital methodology. The impact within the agency appears, to the outsider, to have been substantial.

A further advantage of this participation was the opportunity for immediate user assessment of those areas in which the current technology fell short, thereby pointing the direction for further development.

II. APPLICATION OF DIGITAL PROCESSING

Figure 1 summarizes the data processing and analysis procedures which were applied to the LANDSAT data for developing land use information. More detailed descriptions may be found in Ellefsen et al.[5]

The digital processing procedures proved capable of producing accurate land use classifications over large areas, at large and small scales, with a high degree of detail. Some specifics follow.

Geometric registration of LANDSAT data over the same scene but from contrasting seasons proved to enhance the capability to discriminate urban land use classes from nonurban classes. It was found that at any given time of the year there tended to be unresolvable spectral confusion between urban and non-urban classes. However, the differing rates of change of these classes provided information which improved the discriminability of these "confusion classes."

Geometric correction of the LANDSAT imagery served two purposes. First it made it easier for the data analyst to correlate the LANDSAT data with the available reference data (aerial photography and maps). In addition, after considerable refinement of both the registration and correction techniques based on the needs of this project, it was possible for imagery produced by the classification process to be used directly as a land use map at standard scales such as 1:24,000.

Through judicious use of area sampling and spectral analysis by clustering, it was possible to characterize the ground cover classes in the scene down to Level II, even in the face of relatively limited reference data (primarily high-altitude color and color infrared photography).[6] Often even finer classification detail was available. In most cases involving multitemporal data, feature selection was used to reduce the number of spectral bands used in maximum likelihood classification to between four and six. However, a full 8-channel classification was performed on a set of multitemporal data over the Washington, D.C., area. This was done using the ILLIAC IV parallel processing computer to demonstrate the potential power of the digital processing approach. In roughly eight minutes of processing time, the ILLIAC IV completed a classification (8 channels, 26 classes, 2.5 million pixels) which would have required about 30 hours on the LARS computer (an IBM System/360 Model 67).

The results of the classifications were prepared in both map and tabular formats. Line-printer maps, in black-on-white and most typically at a scale of 1:24,000, provided the most inexpensive form of map output. However, color-coded maps were also obtained at a variety of scales from filmwriting systems by Dicomed, Optronics, and Meade Technology. The results obtainable from such systems are quite variable in cost, quality and versatility.[1] The technology for producing such products continues to evolve rapidly.

The tabular results were produced on a number of bases. Early in the project, UTM grid cells, 1-km and 5-km on a side, were digitized by hand and applied by computer to the classification results. The number of acres and hectares of each land cover class was tabulated for each grid cell. Later, county boundaries and census tract boundaries were encoded from maps using a table digitizer, then geometrically transformed and registered to the LANDSAT data and classification results. Once this was accomplished, the computer could tabulate, by county or census tract, the number of acres or hectares of each land cover class within the area of interest. Still another alternative was to tabulate the percentage of each area classified into each land cover type. Figure 2 shows the census tract boundaries, as digitized and registered, in the Washington, D.C., area. Figure 3 shows a sample of the tabular results which were produced.

These digital processing techniques were applied wholly or in part to LANDSAT scenes of the San Francisco Bay area; Phoenix, Arizona; Springfield, Missouri; Indianapolis, Indiana; and Washington, D. C.

III. ADDITIONAL EXPLORATORY RESEARCH

Subjects of exploratory research included (1) techniques for detecting and mapping change, and (2) use of spatial information, such as texture, to improve classification accuracy.

A. CHANGE DETECTION

The availability of data from multiple satellite passes over a given site plus the technology to precisely register the data has proved valuable in many respects. As noted above, the multitemporal information can be used to increase the classification accuracy. But the dynamics of land use are of interest in themselves, and detection and mapping of land-use change were specific objectives of this project.

In analyzing LANDSAT data from the Phoenix area, Ellefsen et al reported detection of spectral classes indicative of change.[5] To further pursue this matter, a data set was prepared for which several anniversary passes (i.e., separated by a year or multiples of a year) were available together with supporting reference data. The specific objective of this investigation was to determine if land-use changes could be detected by comparison of two classifications of data collected approximately three years apart. The specific land-use change of interest was urban encroachment on agricultural land. The data used was collected over the vicinity of Clermont, Indiana, by the LANDSAT satellite in September 1972 and in October 1975. The basic approach used was to classify each date separately and compare the classifications in such a manner that green or vegetative cover in 1972 which had changed to non-green cover in 1975 would presumably indicate change to urban land use. The data sets were overlaid and precision corrected to a lineprinter scale of 1:24,000. (A third date from September 1973 was included in the data set but was not used in the classifications because little land-use change had taken place in the one-year time interval.) The area was classified with both the standard LARSYS pointwise classifier and the experimental sample classifier (ECHO: see below). The classifications were compared with available photography acquired as close as possible to the dates of data collection. In each case the photography and data were separated in time by less than a year.

The results of this investigation indicated that urban encroachment into agricultural areas can be detected through classification of multitemporal data. However, the size of the change areas must be large relative to the resolution of the sensor, and the classifications must be done with relatively high accuracy with respect to the classes which will show the change. In the part of the midwest from which the data for this investigation were taken, there is almost complete utilization of land for agriculture during the growing season and change becomes relatively easy to find. The analysis has only to detect change from green cover to non-green cover provided the dates have been selected in the part of the growing season when full ground cover is still assured. Land use changes which did not disturb forest cover, such as residential housing underneath partially forested areas, were not detectable using this approach. Since it is not simple change per se which is being detected but rather its multispectral manifestations, the strategy needed to detect and map land-use change is likely to vary with the region involved.

Sample classification is recommended for analysis of data sets between which change is to be detected, since this type of classification tends to eliminate random classification errors and can generally raise the accuracy level toward that needed for reliable detection of land-use change.

B. USE OF SPATIAL INFORMATION

During the first decade of research and development in applying digital analysis techniques to multispectral remote sensing data, emphasis was concentrated on extracting information from the spectral domain. In other words, the methods applied were in the main those which analyze the spectral measurements on a pixel-by-pixel basis. Very little effort was directed at extracting spatial information based on shape, context, texture and other forms of spatial relationships--even though we know from experience with manual interpretation of multispectral remote sensing data that the spatial information content is significant.

There were a number of compelling reasons for looking into the use of spatial features as part of this study. To begin with, we believed that spatial information could be used to improve on the level of classification accuracy attainable using the spectral data alone. Also, while the spectral data is well suited to identifying land cover, the real goal of this study was identification of land use. Thus, while spectral data could be used to classify "asphalt," spatial data might lead us to classifications such as "highway," "roof top," and so on. Of particular interest as we began to consider using spatial information was further improving the distinction between urban cover types and spectrally similar non-urban cover types.

We elected to investigate two approaches to the use of spatial information: sample classification with automatic object finding, and use of scene texture.

ECHO (Extraction and Classification of Homogeneous Objects) is a computer program which analyzes the second order statistics of neighborhoods to "grow" objects and then classifies each object as a whole rather than pixel-by-pixel.[7] Although this is a rather rudimentary use of the spatial information in a scene, it was found to provide significant improvement in urban/non-urban discrimination. Furthermore it proved to be more consistent than the texture measures investigated and was computationally much faster and less expensive to apply. The interested reader is referred to Swain for details.[1] An additional benefit of the ECHO analysis was noted earlier, namely, improvement in change detection results through reduction of random classification errors in the individual classification results.

IV. CONCLUDING REMARKS

We have shown that the synoptic view from satellite altitude together with modern sensor and computer technologies have much to offer those who need accurate and timely land use information. Even with relatively limited "ground truth," computer-assisted analysis of digital multispectral remote sensing data has provided accurate classification of a wide range of ground cover types. From the classification results, map-like imagery and tabular summaries can be produced automatically to serve a variety of land-use mapping and inventorying applications.

Notable about the particular project discussed here was the success with which the evolving technology was transferred from the university research laboratory to the potential user agency. The "user-in-residence" approach not only facilitated the flow of technology but also helped to provide direction for its further evolution.

The possibilities for the future are exciting. Based on research with aircraft and Skylab multispectral data, we know that the sensor systems aboard LANDSAT-C (practically on the launch pad) and LANDSAT-D (still on the drawing board) will further enhance our capabilities to gather remote sensing data for land-use applications. The ILLIAC IV experiment has provided a hint of the power that can still be brought to bear to reduce the data to useful information through automatic and computer-assisted processing.

ACKNOWLEDGEMENTS

Substantial contributions to this research were made by people at LARS, USGS and elsewhere too numerous to mention. The author trusts that thanks enough have accrued to them through the publication of their results and the further successes which have been built on the developments of this investigation.

REFERENCES

1. Swain, P.H., "Land Use Classification and Mapping by Machine-Assisted Analysis of LANDSAT Multispectral Scanner Data," Final Report on U.S. Geological Survey Contract No. 14-08-0001-14725, LARS Information Note 111276, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana, 47907, November 1976.

2. Nichols, J.D., "Pacific Northwest Resources Inventory Demonstration," Proc. Symposium on Machine Processing of Remotely Sensed Data, Purdue University, June 1976, pp. PC-10 to PC-17. IEEE Catalog No. 76CH 1103-1 MPRSD.

3. Ray, R.M., III., and H.F. Huddleston, "Illinois Crop-Acreage Estimation Experiment," Proc. Symposium on Machine Processing of Remotely Sensed Data, Purdue University, June 1976, pp. PB-14 to PB-21. IEEE Catalog No. 76CH 1103-1 MPRSD.

4. Weismiller, R.A., "Land Use Inventory of the Great Lakes Basin Using Computer Analysis of Satellite Data," Final Report on U.S. Environmental Protection Agency Contracts No. 68-01-2551 and 68-01-3552, LARS Information Note 011077, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana, 47907.

5. Ellefsen, R., L. Gaydos, P. Swain, and J.R. Wray, "New Techniques in Mapping Urban Land Use and Monitoring Change for Selected U.S. Metropolitan Areas: An Experiment Employing Computer-Assisted Analysis of ERTS-1 MSS Data," Commission VII Symposium, International Society of Photogrammetry, Banff, Alberta, Canada, October 1974.

6. Anderson, J., E. Hardy, and J. Roach, "A Land-Use Classification System for Use with Remote Sensor Data," U.S. Geological Survey Circular 671, Washington, D.C., 1972.

7. Kettig, R.L., and D.A. Landgrebe, "Computer Classification of Remotely Sensed Multispectral Image Data by Extraction and Classification of Homogeneous Objects," IEEE Trans. Geoscience Electronics, vol. GE-14, pp. 19-26, January 1976.

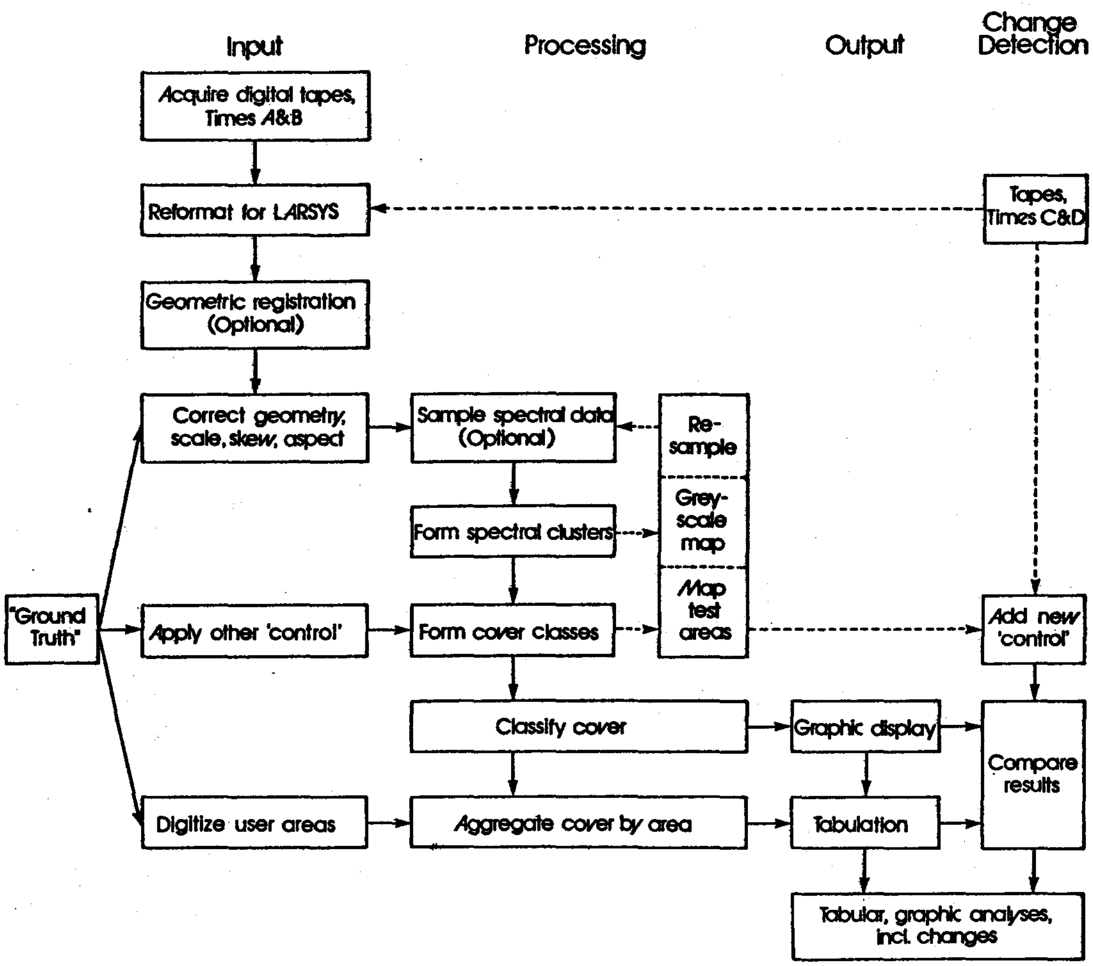

Figure 1. Procedures for Computer-Aided Classification of Land Cover from LANDSAT Multispectral Scanner Data[1]

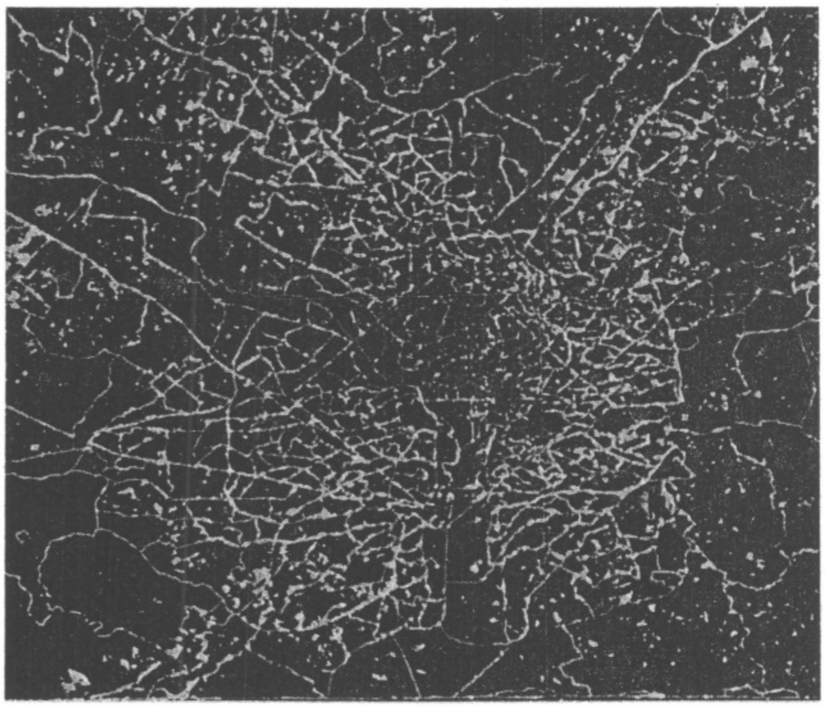

Figure 2. Census Tract Boundaries Superimposed on LANDSAT Imagery of the Washington, D.C. Area.

Number of Pixels Classified Into

Tract No.	Tract ID	No.of Pixels	RSDS	RSDM	DISTRB	CM/IND	PKLOT	OPEN	AGRI	WDLND	TREES	WATER
1	*	886228	87123	2334	9069	280928	1152	4481	126722	37041	189264	[illegible]
2	700500	16706	1910	35	23	47	8	729	7981	2140	3197	
3	700400	5457	494	1	11	12	5	188	2268	766	1[illegible]	[illegible]
4	700301	243	40	0	6	5	0	4	120	32		15
5	700301	44	0	0	0	0	0	0	2		[illegible]	14
6	700300	1529	231	1	30	5	8	34	65[illegible]		724	29
7	700802	3056	429	3	29	7	8	249		[illegible]	243	1
8	700803	1697	381	0	140	34	7	76		105	92	3
9	700700	2298	681	24	44	74	36		[illegible]88	69	16	2
10	700801	3153	606	5	81	33	2[illegible]		1043	161	187	0
11	700800	1457	139	2	18	15		13	622	210	339	1
12	700704	1165	227	3	32	[illegible]	[illegible]	175	2777	480	758	0
13	700703	704	251	5	10		38	1958	10958	2942	6837	895
14	700701	1838	294	2		18	4	7	1090	395	387	2
15	700702	1613	[illegible]		30	19	2	168	744	166	77	0
16	7001[illegible]		[illegible]45	0	7	14	10	8	112	39	127	0
	[illegible]	441	207	5	8	11	2	19	123	28	38	0
[illegible]2	700900	350	161	9	9	56	22	0	61	21	11	0
23	700901	345	168	2	6	43	48	0	50	17	11	0

*Unclassified

Percent of Area Classified Into

Tract No.	Tract ID	No. of Pixels	RSDS	RSDM	DISTRB	CM/IND	PKLOT	OPEN	AGRI	WDLND	TREES	WATER
1	*	886228	9.8	0.3	1.0	31.7	0.1	0.5	14.3	4.2	21.4	15.0
2	700500	16706	11.4	0.2	0.1	0.3	0.0	4.4	47.8	12.8	19.1	3
3	700400	5457	9.1	0.0	0.2	0.2	0.1	3.4	41.6	14.0	31.4	
4	700301	243	16.5	0.0	2.5	2.1	0.0	1.6	49.4	13.2	14.8	
5	700301	44	0.0	0.0	0.0	0.0	0.0	0.0	4.5	34.1	61.4	[illegible]
6	700300	1529	15.1	0.1	2.0	0.3	0.5	2.2	42.7	6.7	[illegible]	0.1
7	700802	3056	14.0	0.1	0.9	0.2	0.3	8.1	43.3	13.[illegible]	[illegible]	0.3
8	700803	1697	22.5	0.0	8.2	2.0	0.4	4.5	3[illegible]		2.3	0.3
9	700700	2298	29.6	1.0	1.9	3.2	1.6	1[illegible]		8.8	10.2	0.0
10	700801	3153	19.2	0.2	2.6	1.0	0.7		[illegible].6	13.0	21.0	0.1
11	700800	1457	9.5	0.1	1.2	1.0	[illegible]	[illegible]	59.8	10.3	16.3	0.0
12	700704	1165	19.5	0.3	2.7	3.[illegible]		7.4	41.6	11.2	26.0	3.4
13	700703	704	35.7	0.7	1.4		[illegible].2	0.3	46.0	16.7	16.3	0.1
14	700701	1838	16.0	0.1		[illegible]	0.1	11.7	51.6	11.5	5.3	0.0
15	700702	1613	24[illegible]			2.1	1.5	1.2	16.9	5.9	19.2	0.0
16	7001[illegible]				1.8	2.5	0.5	4.3	27.9	6.3	8.6	0.0
			[illegible]	2.6	2.6	16.0	6.3	0.0	17.4	6.0	3.1	0.0
	[illegible]	345	48.7	0.6	1.7	12.5	13.9	0.0	14.5	4.9	3.2	0.0
24	700902	848	29.5	0.0	2.8	4.7	4.1	0.9	22.8	6.3	28.5	0.4
25	701100	561	73.8	0.2	0.0	2.0	1.2	0.0	15.5	2.7	4.6	0.0
26	701207	3662	15.7	0.2	1.5	1.1	0.6	5.6	41.8	13.0	18.2	2.2

* Unclassified

Figure 3. Classification Results Tabulated by Census Tract for the Washington, D.C. Area (partial listing).

Philip H. Swain is a professor in the School of Electrical Engineering, Purdue University, and program leader for data processing and analysis research at the University's Laboratory for Applications of Remote Sensing (LARS). B.S. Lehigh University; M.S., Ph.D., Purdue University. Prof. Swain has been affiliated with LARS since its inception in 1966 and has contributed extensively to the development of data processing methods and systems for the management and analysis of remote sensing data. His areas of specialization include theoretical and applied pattern recognition and methods of artificial intelligence.